Vol. 16 No. 10, pp. 273-308 APRIL 1955

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

LIGHT-BEACONS TO AID LANDING AIRCRAFT

by J. B. de BOER.

628.971.8:629.139.1

During the Melbourne air race in 1934, one of the winning aircraft, the "Uiver" of K.L.M., made a remarkable emergency landing at night on an improvised runway at Albury in Australia. On this occasion, improvised lighting was provided by the headlights of motor cars parked side by side along the runway. Compared with such an achievement, landing an aircraft at night on an airfield equipped with modern aids is a relatively simple matter. Modern lighting equipment, which enables aircraft to land safely in darkness and fog, is essential to the smooth operation of the vastly increased volume of civil and military air transport.

About 15 years ago, an aerodrome was usually a flat, grassy field a mile or so square. Pilots were permitted to land or take off from such a field in any direction they chose. Flying took place only when weather and visibility were good, and night landings were accomplished with the aid of lights arranged, say, along the edge of the airfield, on various obstacles, and so on, and a more or less uniform overall illumination of part of the field (about 600 × 800 yds, with an illumination-level of a few tenths of a foot-candle). This illumination was provided by transportable light sources arranged on the boundary of the area giving a fan-shaped light beam in the landing direction along the surface of the airfield [1] [2].

Since that time, the problem of aerodrome lighting as an aid to the landing of aircraft has changed radically for two reasons:

a) Aircraft of far greater weight and speed are flying nowadays, and concrete runways from 1000 to more than 2500 yds long and 100 to 200 ft wide are required to enable such aircraft to take off and land in complete safety. It is impossible to provide adequately uniform overall illumination over such an area.

b) It is the aim nowadays to make air-transport less and less dependent upon weather-conditions; hence aircraft must be enabled to land not only day or night when visibility is good but also in bad visibility at night and — most difficult of all — under conditions of mist or fog in the daytime. In these latter circumstances it is impossible to make objects other than light sources visibile at a sufficient distance.

These two facts have led, of necessity, to the replacement of *lighting* of the landing area, by *beaconing*. In fact, beaconing is the basis of all up-to-date systems of "airfield lighting".

However, also the light sources that can reasonably be used for this purpose, are visible in heavy mist at a distance of one or two thousand feet at the utmost. Such a short distance gives the pilot of an approaching aircraft no margin for important manoeuvres; landing in heavy mist has therefore become possible only with the help of a combination of light-beacon and radio-beacon systems (or radar information conveyed to the pilot by radio).

The radio beacons enable the pilot to to be guided so close to the airfield that he can just discern the lights, after which the pilot lands the aircraft visually with the aid of the lights, since radio beacons are not sufficiently accurate at short range.

The bad-visibility landing procedure may now be described more fully. Radio or radar is used to guide the aircraft as accurately as possible along the ideal glide path, that is, a line in the same vertical plane as the centre-line of the runway, sloping down at an angle of about $2^1/_2^\circ$ and cutting the above-mentioned centre-line at a point about 1000 ft beyond the start (or "threshold") of the runway. By the time the aircraft has descended to a height of about 200 ft (at which the horizontal distance to the threshold of the runway is still about 3000 ft), visual "ground-contact" should be established. From this contact the pilot must obtain an immediate and unmistakable impression of a number of data: his direction of flight relative to the direction of the runway; the horizon; the altitude of the aircraft; the glide slope; the distance to the threshold; the lateral displacement from the centre-line of the runway (produced) and the ground speed of the aircraft. Experience has shown that it is possible to provide the pilot with all this information by the mere observation of a number of light-beacons provided these beacons are arranged in a suitable pattern, and show if necessary, distinctive colours. On the basis of this information obtained visually the pilot performs all further manoevres for completing the landing.

---

[1] G. L. van Heel, The illumination and beaconing of aerodromes, Philips techn. Rev. 4, 13-99, 1939.

[2] Th. J. J. A. Manders, Aerodrome illumination by means of water-cooled mercury lamps, Philips tech. Rev. 6, 33-38, 1941.

It will be evident from the above that the lights employed as visual guides do not have to be visible from all directions; fortunately, then, light sources of reasonable power can be used by beaming the luminous flux in the required direction. Again, the restriction of the light-yield to certain directions prevents excessive diffusion of the light in heavy mist, which would make the pattern of the lights difficult to distinguish from an approaching aircraft.

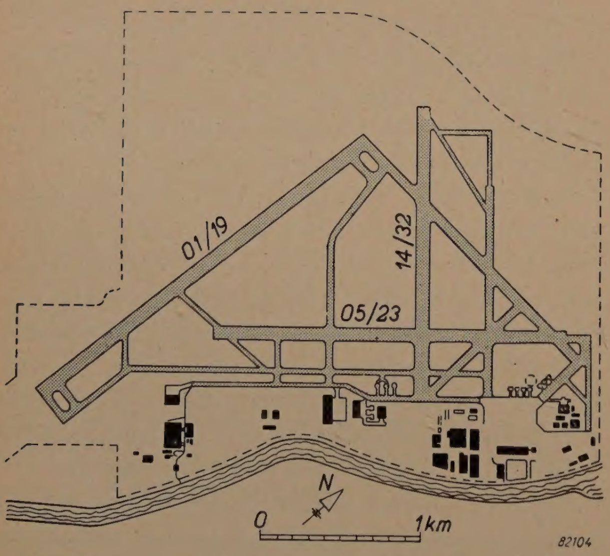


Fig. 1. Plan showing the present layout of Schiphol airport, Amsterdam. Each runway offers two possible directions for landing or take-off, depending on the wind, and must therefore be provided with approach and threshold lighting at both ends, and runway lights shining in both directions. In view of the directions of the prevailing winds, Schiphol is now equipped with two main runways, i.e. runway 05/23 (take-off and landing directions 50° or 230° to the S-N direction) and runway 01/19 (10° or 190° to the S-N direction). The beacon systems on these runways are formed by lights designed by Philips. (This plan, the diagram of fig. 3 and photographs of fig. 4, fig. 14 and fig. 21, is reproduced here by permission of the Schiphol Airport Authorities.

In this article it will be shown how the required luminous intensity and light-distribution of the beacons can be determined. The construction of some of the lights will then be described, with special reference to the lighting actually in use at a number of airports, e.g. on the runways at Schiphol Airport, Amsterdam (*fig. 1*).

For the sake of clarity, however, it is first necessary to deal more fully with how the pilot obtains visual guidance during a landing.

Configuration of the lights

Since the second world war, many experiments have been carried out on airfields, and in the laboratory, with the object of determining the most suitable pattern for runway beacons [3]). That the runway itself must be indicated by placing rows of lights along the sides and at the threshold is self-evident, but a certain amount of controversy has arisen in connection with the so-called approach lighting, that is, the lighting to indicate the position of the runway before the latter is actually seen through mist or fog. When once the IATA (International Air Transport Association) had decided its standpoint with regard to this question, the ICAO (International Civil Aviation Organisation) meeting in Montreal in November 1952 made a fairly clear choice between the various systems of approach lighting which had been proposed. *Fig. 2* shows one of the systems consistent with the recommendations of this organization, namely the Calvert system. Here, the extension of the centre-line of the runway is indicated by a luminous line, at right-angles to which, at specified intervals, are luminous transverse lines ("cross-bars"), arranged in order of diminishing length towards the runway. Each of these luminous lines really consists of a number of lights; hence the lines are usually seen as straight rows of luminous dots rather than as continuous lines.

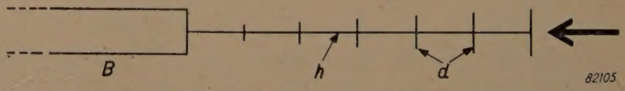


Fig. 2. Diagram showing the Calvert system of runway beaconing; the thick arrow indicates the landing direction. *B* runway, *h* centre-line, *d* cross-bars.

---

[3]) E. S. Calvert, J. Roy. Aeronaut. Soc. **52**, 439, 1948; E. S. Calvert, Trans Ill. Eng. Soc. (London) **15**, 183, 1950; H. J. Cory Pearson, C. A. A. Techn. Dev. Rep. Nos. 104 and 167, March 1950 and April 1952; A. N. Baldino, Bull. Soc. Trans. Electr. (6) **9**, 442, 1949; G. J. Malouin, Report to Flight-Technical Group IATA, New York, October 1951; J. W. Sparke and H. F. Ringe, Light and Lighting **43**, 259, 1950; F. C. Breckenridge, Illum. Eng. **47**, 1952.

A variation of this system, as shown in *fig. 3*, is employed on main runway 23 at Schiphol. *Fig. 4* is a photograph of this beacon system as seen from the cockpit of an approaching aircraft.

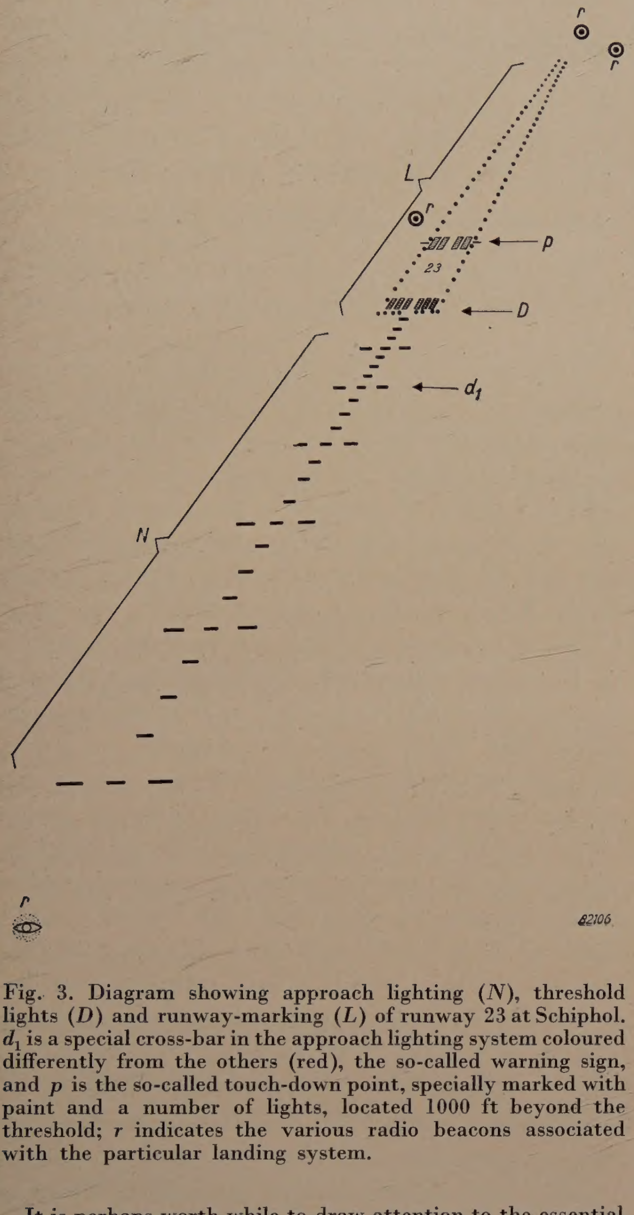


Fig. 3. Diagram showing approach lighting (N), threshold lights (D) and runway-marking (L) of runway 23 at Schiphol. d_1 is a special cross-bar in the approach lighting system coloured differently from the others (red), the so-called warning sign, and p is the so-called touch-down point, specially marked with paint and a number of lights, located 1000 ft beyond the threshold; r indicates the various radio beacons associated with the particular landing system.

It is perhaps worth while to draw attention to the essential purpose of the cross-bars in the pattern. When an aircraft is approaching the runway along the "ideal" glide path and in the absence of cross-wind, the pilot will see the centre-line of the pattern as a vertical line straight ahead of him (a pilot, like a motorist can judge from experience what point corresponds to the direction "straight ahead" from its relation to the edge of his windscreen or the frame of his machine. In these circumstances, the pilot will be able to judge the distance to the threshold correctly if a suitable code is incorporated in the centre-line (e.g. differences in the colour or grouping of the individual lights). Assuming once more that there is no cross-wind, then, if the pilot sees the centre-line in the perspective pattern in another position, this may indicate: *a*) that the longitudinal axis of the aircraft is not parallel to the centre-line; *b*) that the aircraft is flying wide of the centre-line; or *c*) that the line joining the wing-tips is not horizontal. To illustrate these ambiguities, *fig. 5* shows two different situations in which the centre-line will be seen by the pilot as running in an identical direction. In the case shown in fig. 5*a*, the aircraft

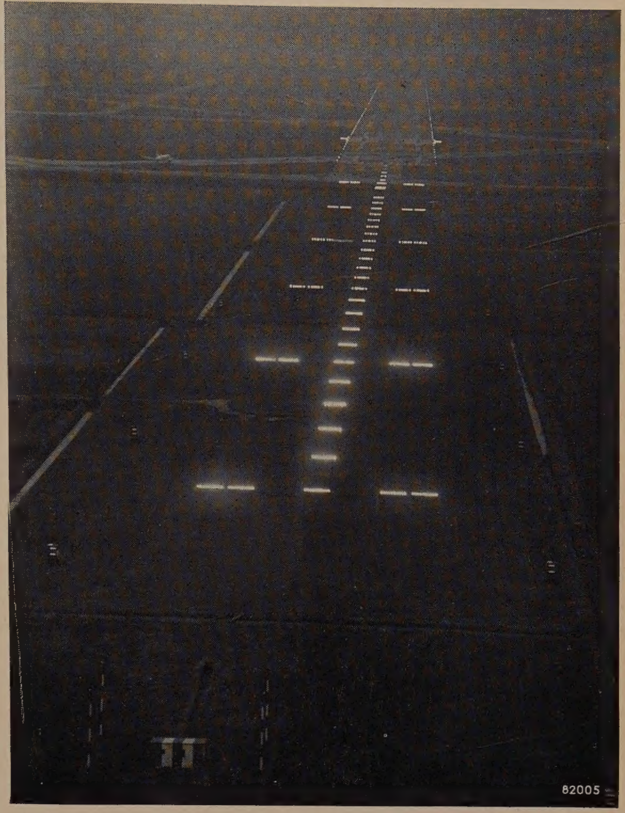


Fig. 4. Beacon system of Schiphol runway 23 as seen from an aircraft coming in to land.

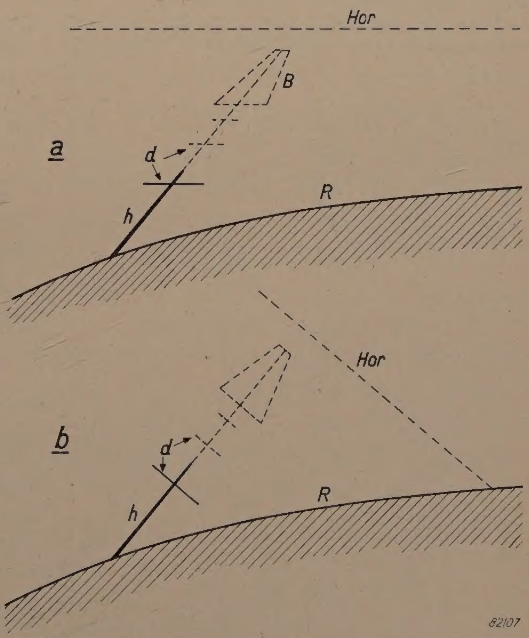


Fig. 5. Diagram to illustrate two different situations in which the pilot of an approaching aircraft will see the centre-line (h) of a runway (B) in exactly the same position relative to his cockpit cut-off. The horizon (Hor) is assumed to be completely obscured by fog in both cases. It is only by observing the attitude of the cross-bars (d) that the pilot is able to judge his position and direction of movement correctly.

is flying with the line joining its wing-tips horizontal, on a course parallel to, but slightly to the right of the runway: fig. 5*b* shows the situation when the aircraft is above the approach centre-line of the runway with its fuselage momentarily parallel to this line, but is banking to the left, so that the line joining the wing-tips slopes downwards to the left. It is only by observing the relative attitudes of the cross-bars that the pilot can judge which of these two interpretations corresponds to the real situation and so choose the proper manoeuvre to correct his course.

In view of the speed at which the aircraft travels, it is an interesting psychological problem as to how the pilot is able to deduce his position and movement in space from the movements of the different visible points and lines within his field of view. Some light is shed on this question by Calvert's so-called parafoveal streamer theory[3]). In view of the importance of this theory as a means of establishing present-day requirements for the light-distribution of runway beacons, Calvert's arguments will now be outlined.

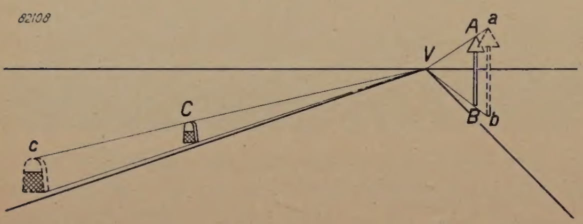


Fig. 6. Diagram showing the field of view of a motorist driving along a straight road. *V* is the vanishing point. Note the traffic sign (*AB*) and the mile post (*C*) as seen by the driver at a particular moment, and the same two objects (*ab* and *c*) as seen a moment later. From the apparent motion of all points outwards from the vanishing point, the driver is able to judge his own direction and speed.

Fig. 6 is a simplified diagrammatic representation of a perspective image as seen by a rapidly moving observer, e.g. the driver of a car travelling along a straight road. All the points within the field of view of such an observer appear to move outwards from a single fixed point, that is, the vanishing point (*V*), the point on the horizon corresponding to the infinitely distant "end" of the road). The speed of movement of a given point is proportional to the square of its apparent distance from *V*, and inversely proportional to the distance of this point from the line along which the observer's eye is moving. The more or less instinctive observations of the driver (or pilot) as to the direction and speed of movement of such points (in what may be described as the "dynamic field of view") enable him to estimate the direction and speed of his own movement; the less he alters his direction of view, the more accurate will his estimate be. In fact, films taken by Capt. Majendie of BOAC have shown that during the last phase of a landing, a pilot habitually "fixes" his head and eyeballs, that is, stares ahead at a point corresponding roughly to the point on the runway where the wheels of the aircraft will touch the ground. His line of sight then makes a small angle with the horizon: the zone visible to the pilot between the horizontal and about 10° below it is therefore the most important. Hence the light radiated by the beacons within the angle between 0° and 10° above the horizon contributes most to the visual guidance during landing.

Minimum requirements for visual guidance

The requirements to be imposed on approach systems and runway lights will now be defined more fully.

When once the pilot has established "ground-contact" (at a distance of 3-4 thousand feet from the runway-threshold), he must be able to see at any given moment a certain length of the system of lights in order to appreciate the overall pattern. This is necessary not only for the ideal glide path but also for other approach-tracks, within certain limits. These limits, or more precisely, the space within which visual guidance is to be assured, cannot be determined objectively. The information required to establish it can be supplied only by experienced pilots from their practical knowledge of flying. Recommendations concerning this zone were issued in September 1952 by the Flight-Technical Group of the I.A.T.A., an association well qualified to advise on such matters, since its members include many experienced pilots employed by the world's major air-lines. According to these recommendations, the space within which visual guidance must be provided may be imagined as a shaft of oval cross-section around the ideal glide path; the dimensions of this approach channel are indicated in *fig. 7*. An ordinary commercial aircraft approaching the runway anywhere outside this channel cannot readily, if at all, be so manouvred as to complete a safe landing; hence visual guidance outside the channel is of little value. On the other hand, standard radio or radar aids are quite accurate enough to "home" an aircraft to the "entrance" of the visual guidance channel defined by fig. 7.

Originally, it was considered necessary to require that a segment of constant length of the beacon system be visible from all points within the guidance channel. The light just visible above the bottom edge of the windscreen, the so-called cockpit cut-off (about 15° or 20° below the horizon, *fig. 8*), is a natural choice for the near limit of the visible segment. The consequence of the above consideration would be that the maximum visual range (or luminous intensity) of the *approach* lights must be in quite a steep upward direction, and the maximum luminous intensity required of the *runway* lights may decrease at increasing distance from the runway-threshold.

It will be evident that such a system is altogether inconsistent with the parafoveal streamer theory, which holds that it is especially necessary that the pilot be able to see the lights within the angular zone between 0° and 10° below this horizon: this also includes the runway lights far ahead, whose luminous intensity would clearly have to be considerably higher than those of the approach lights and the lights at shorter distances.

Since the fulfilment of the above requirement would lead to enormous luminous intensities (see below), a compromise has been adopted: it is now required that all the lights become visible to the pilot at the same distance, i.e. 1500 feet. No light need be radiated at angles greater than 12-15° above the horizon except by the lights in the near part of the approach system, which must become visible to the pilot before he has "fixed" his eyes.

It will be seen that the required luminous intensity data of each individual light in all directions of radiation can be deduced from the above data and with the aid of the diagram shown in fig. 7, provided that the minimum conditions of visibility in which aircraft must be able to land are known. These conditions cover *a*) the background-brightness with which the lights must contrast and *b*) the transmission of the atmosphere. The luminous intensity to give the lights the required range will be considerably lower for night landings in good visibility than for daytime landings in fog. The I.A.T.A. recommends that a daytime landing in a meteorological visibility of only 1000 ft be considered the most unfavourable case in which full visual guidance should be provided. "Visibility" as a quantitative measure of the transmission of the atmosphere may be defined as the greatest distance (V) at which a dark object can be distinguished by day. The relationship between V and the transmission (r) of the atmosphere is defined in the formula of Koschmieder: $t^V = c$, where c is the contrast-sensitivity of the human eye. For the present type of observations, c may be taken equal to 0.04.

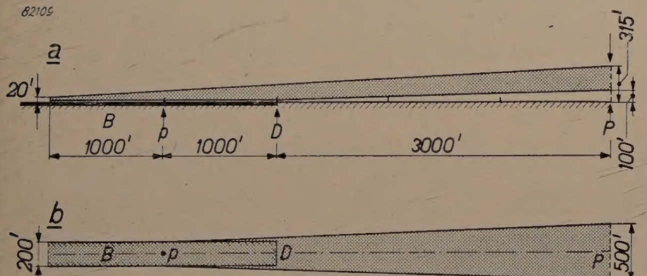


Fig. 7. *a*) Side view, and *b*) top view, of the imaginary shaft within which visual guidance should be given to an aircraft approaching for a landing. *B* runway, *D* threshold, *p* touch-down point, and *P* entrance of the visual guidance channel.

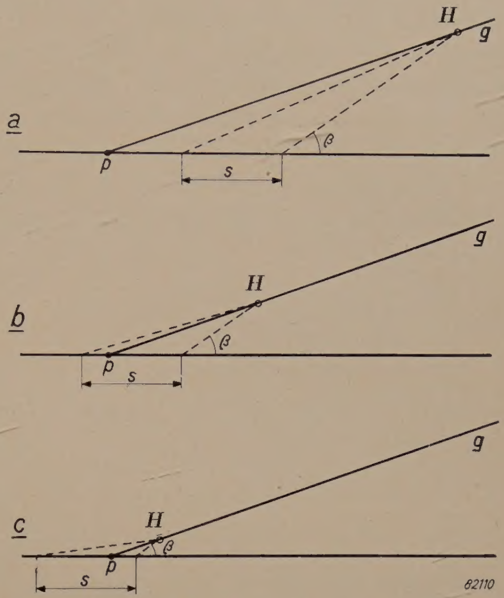


Fig. 8. In order to satisfy the requirement that the length (s) of the beacon system visible from an aircraft (H) approaching along the ideal glide path (g) be the same at all points along this path, the visual range of the initial approach lights would have to be longer than that of the others, in directions slanting quite steeply upwards. β is the limiting angle at which the pilot of an approaching aircraft will just be able to see a particular light above his cockpit cut-off. To illustrate the point more clearly, the glide path is here drawn at an exaggerated angle.

For reasons of economy, only those airfield runways corresponding to wind-directions frequently associated with relatively dense mist will be provided with approach lighting for $V = 300$ m. At Amsterdam, for example, a north-east wind very rarely brings fog; hence the two principal runways (01/19 and 05/23, see fig. 1) are provided with strong approach lights only at the north and north-east ends, respectively, the lighting at the south and south-west ends of these runways, where aircraft land in a prevailing north-east wind, being considerably weaker. The north-west to south-east runway (14/23 in fig. 1) is provided with relatively weak approach lighting at both ends.

Determination of the required luminous intensity and light-distribution of the beacons

The relationship between the visual range of r (in metres) and the luminous intensity I (in candelas) of a light source is defined by Allard's formula:

$$I = \frac{E_c \cdot r^2}{t^r},$$

where t is the transmission of the atmosphere per metre and E_c the minimum eye-illumination (in lux) to make the light visible to the observer. Apart from the background-luminance, the required eye-illumination E_c also depends on several other factors, e.g. the apparent size, the form and the colour of the particular light source [4]. For night-flying, 10^{-6} lux is a suitable value of E_c, and for day-time flying 10^{-3} lux.

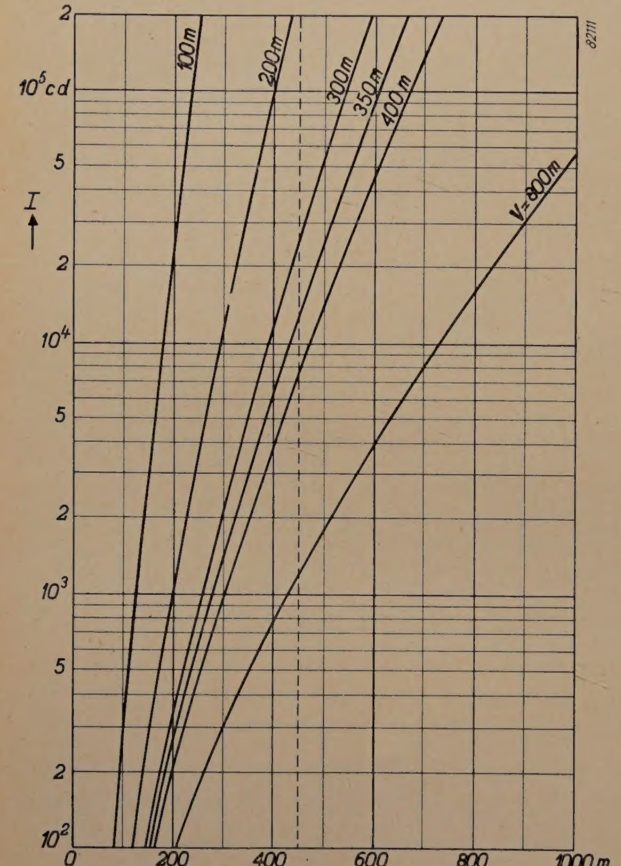


Fig. 9. Diagram showing the relationship between luminous intensity (I, in candelas) and visual range (r, in metres) for different meteorological visibilities (V, in metres). The curves are valid for an eye-illumination level $E_c = 10^{-3}$ lux.

Fig. 9, plotted for $E_c = 10^{-3}$ lux, shows that the effect of meteorological visibility on the luminous intensities required to ensure the prescribed visual range is considerable. It may easily be deduced from this diagram that any tightening of the requirements as to the range at which the lights must be visible or the maximum density of fog in which a landing may be attempted (V) will involve considerable expense. For a visibility $V = 400$ m (1300 ft), a light of 4000 cd luminous intensity can be seen at a distance of 400 m; to increase the range to 600 m (2000 ft) it, would be necessary to raise the luminous intensity of the light source to 45 000 cd. Again, in a visibility V of only 200 m, (670 ft) a range of 400 m will require a luminous intensity of 100 000 cd, and a range of 600 m some million cd. To appreciate the significance of these figures it should, be borne in mind that whereas a luminous intensity of the last-mentioned order is not unusual for light-houses, of which there are only fifteen along the entire coast of the Netherlands, it is altogether out of the question for runway lights, of which several hundreds are required on a single runway.

At $V = 300$ m, already referred to as the visibility upon which our calculations should be based, a luminous intensity of 13 000 cd is required for a range of 400 m in daylight; this would have to be raised to more than 200 000 cd to increase the range to 600 m. Hence the above-mentioned decision to specify a range which is not unduly long, that is, 450 m, the corresponding luminous intensity being 25 000 cd.

Let us now consider the required light-distribution.

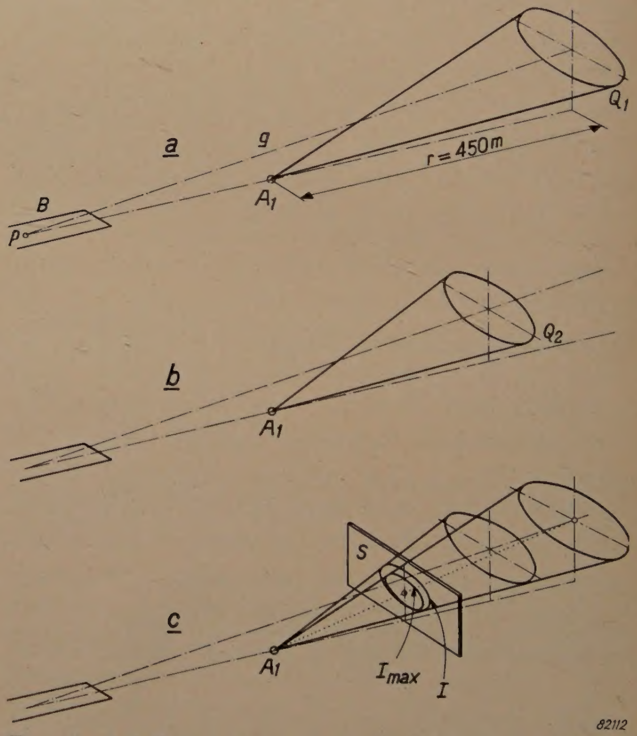


Fig. 10. a) A light (A_1) at a distance $r = 450$ m (1500 ft) from a particular cross-section (Q_1) of the visual guidance channel should possess a luminous intensity $I_{max} = 25\,000$ candelas in the directions of all the points within that cross-section (the small differences in distance of these points from the light may be ignored).
b) A lower luminous intensity (I) is sufficient in the directions of points within another cross-section of the channel (Q_2), closer to A_1.
c) Combination of diagrams (a) and (b), by projection on an imaginary screen (S), forming an isocandela diagram which defines the required light-distribution of the light A_1.

---

[4] J. B. de Boer, Visibility of approach and runway lights, Philips Research Reports **6**, 224-239, 1951.

An idea of this may be obtained from *fig. 10*. Fig. 10*a* shows a vertical cross-section of the region of visual guidance (fig. 7), taken at a point 450 m from one of the light sources A_1 of the approach lighting system. Since A_1 must be just visible from every point within this cross-section, its luminous intensity must be at least 25 000 cd everywhere within the solid angle subtended by the cross-section at Q_1. Fig. 10*b* shows another cross-section closer to the light source. A lower luminous intensity (in accordance with fig. 9) will be sufficient for directions within the solid angle or cone associated with this cross-section. Of course, the luminous intensity specified for the directions within the cone of fig. 10*a* must be maintained. In fig. 10*c* diagrams *a* and *b* are combined to show the two cones of fig. 10*a* and fig. 10*b* and their intersection with an imaginary surface at right-angles to the extended centre-line of the runway. The resultant curves are "isocandela lines", since they define the limits within which a given luminous intensity, in candela, is required. Accordingly, a complete isocandela diagram fully describing the required light-distribution is obtained by considering a series of cross-sections of the channel of guidance, one after another. Three such diagrams, referring to lights located at three different points are shown in *fig. 11*.

Fig. 10, shows the way in which an isocandela diagram for a light on the centre-line of the system is constructed. The light-distribution of a light displaced laterally with respect to the centre-line, e.g. on a cross-bar, will be virtually the same, but in this case the axis of the beam should be slightly toed-in instead of running parallel to the centre-line. Space does not permit us to amplify this brief explanation. A more precise description of the calculation of the required light-distributions may be found elsewhere [5]).

Given a light having the exact distribution required, it will be just visible to the pilot of an aircraft entering near the "ceiling" of the visual guidance channel; flying lower, however, he will see the light at a luminous intensity greater than the minimum required to make it visible to him. In fact, the pilot of an aircraft flying nearly at the "floor" of the guidance channel will be exposed to the full luminous intensity corresponding to a visual range of 450 m, until he has nearly reached the particular light source. It will be seen that this involves a risk of dazzle, which is all the greater owing to the fact that it is, of course, impossible to design a light whose luminous intensity is exactly the same *on* the inner isocandela line as *within* it; the luminous intensity will inevitably be greater inside the curve, reaching a maximum at the "axis" of the light-beam.

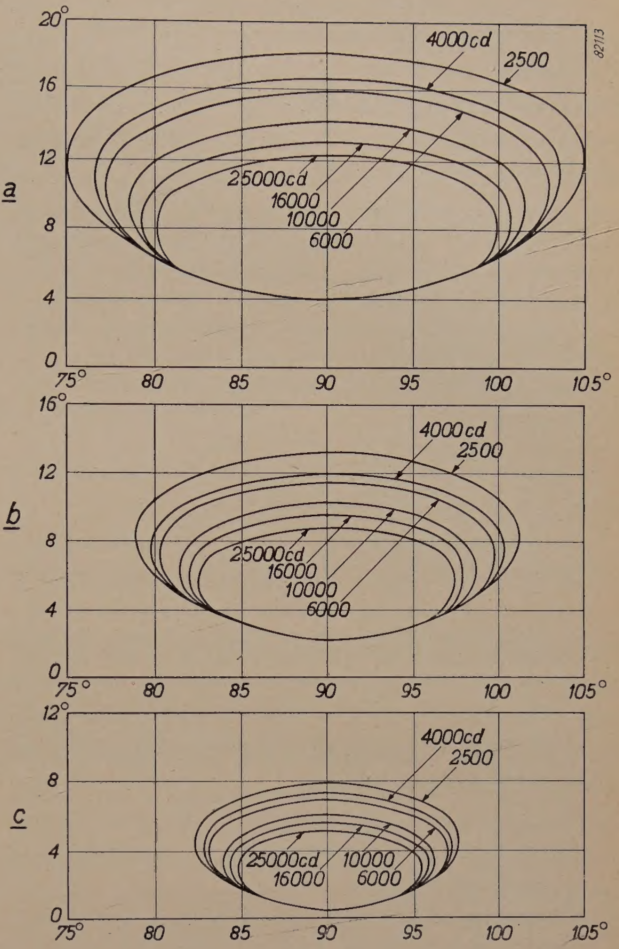


Fig. 11. Diagrams showing the required light-distribution (isocandela diagrams) of *a*) initial approach light, *b*) a threshold light, and *c*) a runway light 450 metres (1500 ft) beyond the threshold.

To avoid dazzle, it is necessary to ensure that the luminous intensity does not exceed a certain maximum value. The *a priori* possibility of satisfying both equally important requirements viz. visibility of the light at the specified range and avoidance of glare at the shortest distance at which the light disappears under the cockpit (angle β in fig. 8) if the pilot follows the least favourable track approach (near the "floor" of the channel), is very much open to question. Fortunately, investigations have shown [6]) that the latitude between the minimum eye-illumination necessary for discerning a light and the maximum eye-illumination tolerable in view of glase, both measured, of course, under the same conditions) is quite large. This latitude may be taken as a factor of 200 at night, and about 1000 by day. The geometrical dimensions of the visual guidance channel and those of the approach lighting system are such that, given an ideal light-distribution as shown in fig. 11 and a visibility not less than 300 m, the above-mentioned latitude will not be exceeded by the luminous intensity anywhere within the channel.

---

[5]) J. B. de Boer, Calculations on the light distribution of approach and runway lights, Philips Research Reports **6**, 241-250, 1951.

[6]) G. A. W. Rutgers, Visibility of signal lights, Electrotechniek **26**, 36-40, 1948 (In Dutch).

It will be evident that for a day landing in relatively better visibility, as for a night landing, a lower luminous intensity of all the lamps is not only permissible, but essential for the avoidance of glare. Hence the overall luminous intensity is controlled from a central point by varying the supply current. In fact, one of the methods employed enables the pilot of an approaching aircraft to vary the luminous intensity of the beacon-system to suit his own immediate requirements by means of a radio-operated servo-mechanism.

The supply system of an airfield beacon system will be described in a separate article in this Review.

The practical realization of the calculated light-distributions involves its own special problems. For reasons of economy it will be necessary to obtain the required light-concentration with the aid of fairly simple optical systems; hence a reasonable approximation to the ideal light distributions is the most that can be expected in practice. The principal aims of the designer will be:

1) To reduce the above-mentioned peak luminous-intensity in the beams as far as possible, and so limit glare.
2) To minimize the amount of light radiated outside the prescribed beam, since the light so "spilled" not only constitutes a loss in itself, but, owing to diffusion in fog or mist, increases the overall background luminance and so necessitates a higher level of eye-illumination: i.e. the range associated with a particular luminous intensity is reduced. For the same reason, it is necessary to ensure that the amount of light radiated upwards is no larger than strictly necessary.

Design of the different lights

The three types of light constituting the beacon system of a runway, i.e. approach lights, threshold lights and the actual runway lights (see fig. 3), necessarily differ considerably in design, firstly by reason of the difference in the required light-distribution (see fig. 11) and secondly — a very important factor in the design — because they vary as regards the extent to which they may be raised above ground level. The initial approach lights, stationed at a point which approaching aircraft will clear by a hundred feet or so, may safely be raised some feet above the ground (the "rough"). On the other hand, the lights at the threshold of the runway, where aircraft frequently touch down, should preferably be virtually or entirely "flush". The lights at the edges of the runway may protrude above the rough, but, since there is a possibility of the aircraft colliding with them, must then be so designed as to ensure that such a collision cannot damage the aircraft.

Another important feature of design, common to all the lights, is adjustability as regards azimuth and elevation (aiming adjustments). This is necessary as a means of ensuring correct coverage of the visual guidance channel with the particular light-distribution provided (fig. 7). Such aiming necessitates special features of the lamps as well as certain auxiliary equipment, some examples of which will be described briefly at the end of this article.

Incandescent lamps are employed for all the lights considered, since the concentrated beam of light indicated by the diagrams in fig. 11 would be virtually unattainable from relatively large light sources of low luminance, such as fluorescent, or sodium lamps. High-pressure mercury vapour lamps would be suitable by virtue of their very high luminance-level, but are less simple to operate (e.g. it is more difficult to vary their light output) than incandescent lamps; hence they have not been employed hitherto for this purpose.

The various components of the system will now be described more fully, viz. the approach lights, the runway lights and finally intermediate the threshold lights.

Approach lights

The approach lights require a greater vertical and lateral beam-spread than either of the other two classes of light referred to in the above (fig. 11). However, they may be mounted entirely elevated, and need not be unduly restricted as to size and weight.

A special incandescent lamp (see *fig. 12*) has been designed to furnish the required light-distribution. The rear bulb-wall of this is silvered on the inside and has the shape of a paraboloid of revolution, provided with vertical ridges, whose form governs the horizontal spread of the beam emitted by the lamp. *Fig. 13* shows the isocandela diagram of this lamp, which is fitted with a 400 watt low-voltage filament (24 V). Comparing this diagram with fig. 11*a*, we see that the lamp amply satisfies the requirements imposed. Although it is not within the scope of this article to enter into the reasons for choosing such a low operating voltage (or, rather, a relatively strong current) which necessitates a separate transformer for each lamp, it may be noted that a low operating voltage is consistent with a thick, compact filament, which facilitates the concentration of the light.

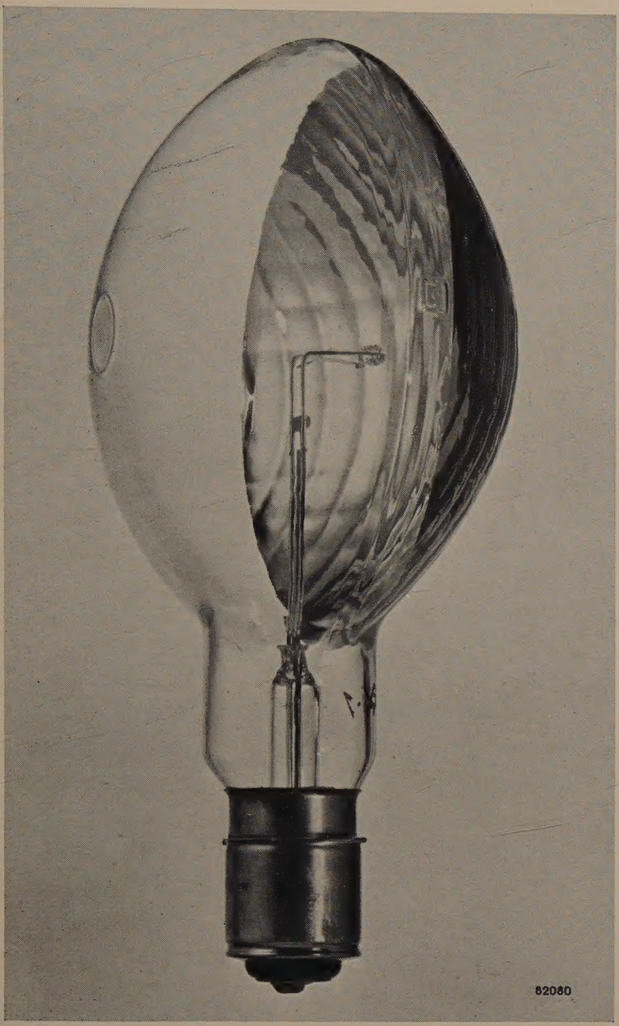


Fig. 12. Special 400 W, 24 V incandescent lamp for use in approach lights. Part of the bulb has the form of a paraboloid of revolution, silvered on the inside and vertically ridged. Height of lamp: 11 inches.

The lamp is fitted with a so-called "prefocus" cap (type P40) having two dissimilar sector-shaped lugs by virtue of which the lamp can be placed in only one precisely defined position in the lamp holder. The orientation of the beam relative to the cap is accurately adjustedand fixed during the manufacture of the lamp, so that correct beam-alignment is assured whenever a new lamp is inserted in one of the lamp-holders which are aimed once and for all when they are installed.

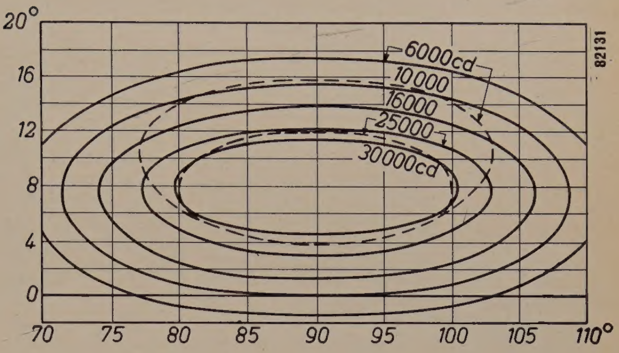


Fig. 13. Isocandela diagram of the 400 W lamp shown in fig. 12. Two curves from fig. 11*a*, drawn as dotted lines, are also included in this diagram.

The bulbs of these lamps are made of hard glass and therefore require no protection from the weather. The principal advantage of such special lamps is that they require no fitting other than an adjustable lamp-holder suitable for outdoor use: hence all difficulties arising from the infiltration of dust and dirt into optical systems and the weathering of mirrors are avoided.

The cross-bars of the approach-lighting system are built-up by arranging several lamps in a row. On one of the Schiphol runways this is accomplished by the method illustrated in *fig. 14*. Here, such a bar comprises a number of units of 10 lamps mounted side-by-side on a bracket, the advantage being that all 10 lamps can be adjusted simultaneously to

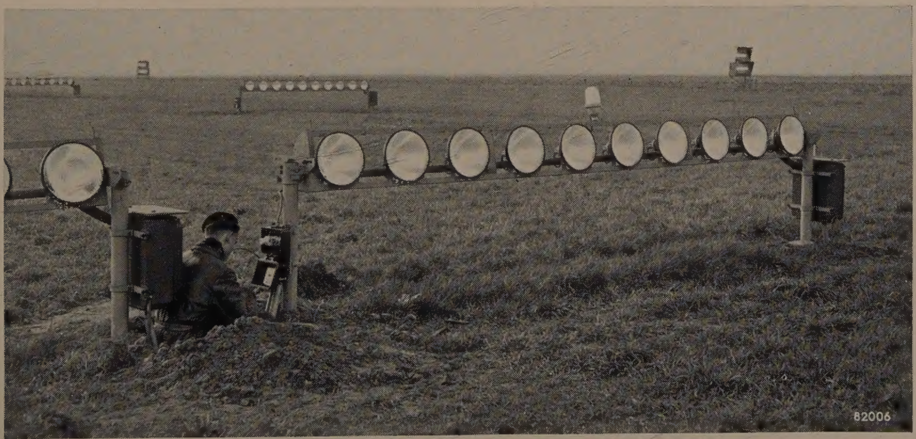


Fig. 14. Bracket carrying 10 approach lights (of a type other than that shown in fig. 12). Several of these brackets make up one "cross-bar" in the approach lighting system of runway 23 at Schiphol.

the correct beam-elevation merely by turning the bracket. A small omni-directional light (that is, an ordinary incandescent lamp in a barrel-shaped Fresnel lens of pressed glass) is also provided at the centre of each bracket. Provided that the visibility is not unduly bad, these omni-directional lights enable the pilots of aircraft circling the airfield to maintain their sense of direction relative to the runway.

Runway lights

As we have already seen, the runway lights may protrude above the level of the rough, if so designed as to cause no damage to any aircraft happening to collide with them. Such raised lights have a number of advanatages over "flush" lights. They not only enable the requirements as to luminous intensity and beam-spread to be satisfied far more easily, but (if suitably designed) are less likely to be covered by snow in winter.

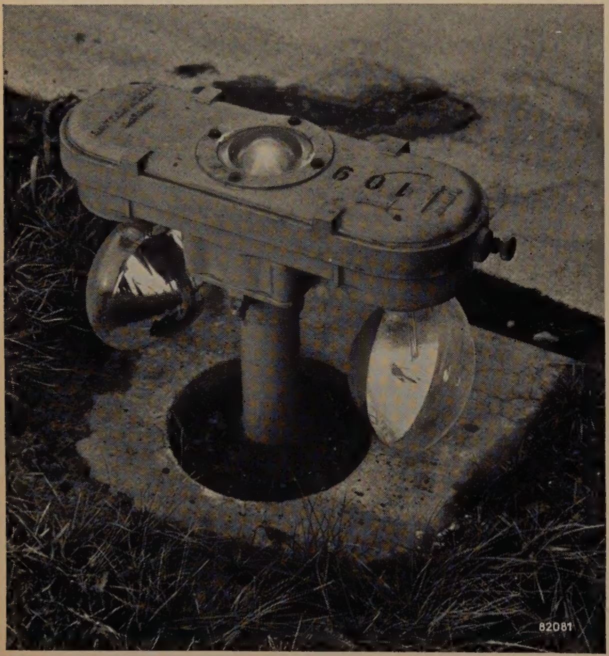


Fig. 15. Elevated runway light, as employed on a number of airfields in the Netherlands. Two special lamps, similar to that shown in fig. 12 but rated at only 100 W, are suspended back to back to shine in the two landing directions of the runway. The cover contains an omni-directional light.

Raised runway lights are employed on several airfields in the Netherlands. *Fig. 15* shows the construction of such a light. It includes two adjustable lamp-holders from which two special lamps are suspended. These lamps have bulbs similar to that shown in fig. 12 (though slightly smaller), and the same type of prefocussing lamp cap, but the filament-power is only 100 W. The light-distribution in the beam of each individual lamp (*fig. 16*) agrees reasonably with the requirement as defined in fig. 11*c*. The two lamps throw their beams in opposite directions, that is, one in each direction from which aircraft may land on the runway. Like the approach unit, the runway light also includes a small omni-directional lamp (35 W).

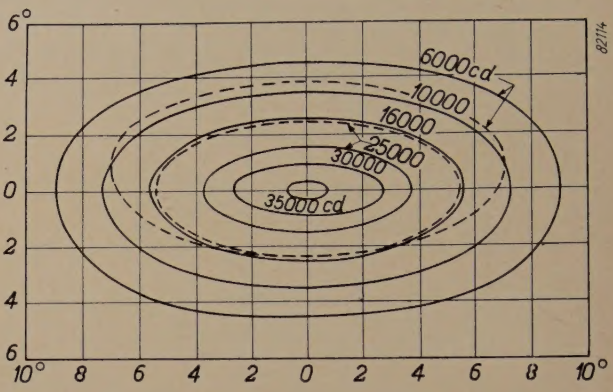


Fig. 16. Isocandela diagram of the 100 W lamp shown in fig. 15, including two curves (dotted lines) from fig. 11c.

These runway lights are provided with so-called breakable coupling, that is, specially weakened joints in the supports at a point just above ground level, which break readily if the fitting is struck by an aircraft. Within these mechanical joints are breakable circuit connections, i.e. plug-socket connections each with a rubber socket in the lead between the lamps and the buried transformers.

Fig. 17 shows a still simpler solution to the problem, involving the use of separate lights for the

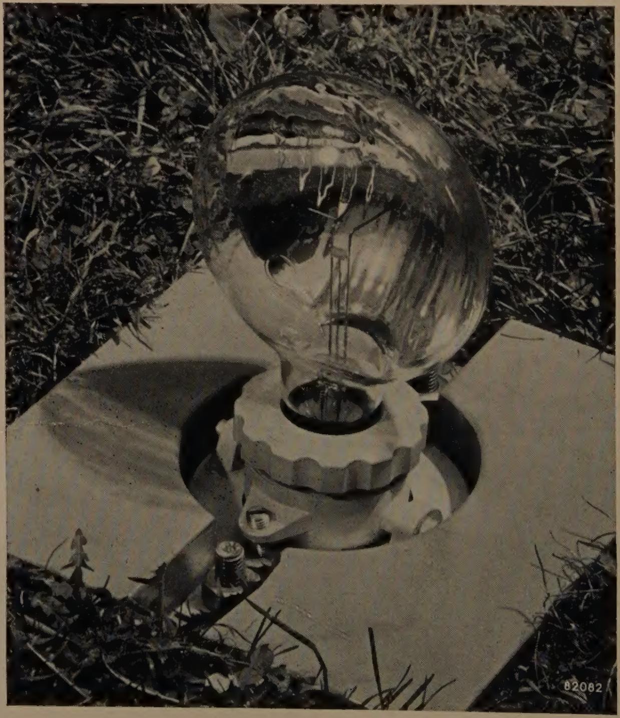


Fig. 17. The same lamp as that shown in fig. 15, mounted in a single adjustable fitting to give a very simple runway light.

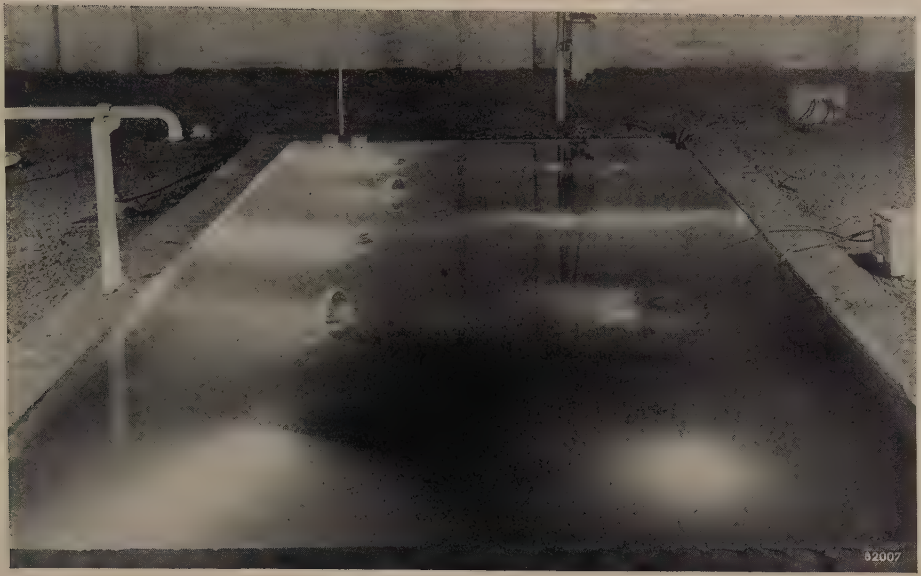


Fig. 18. Testing tank for outdoor lighting fittings. Flush (or semi-flush) lights are mounted as in practice in the tank. The tank is periodically flooded with water and emptied. Five lamps of the type shown in fig. 17 can be seen immersed or semi-immersed on the left-hand side of the picture.

two landing-directions. Here, the special lamp already described is fitted upright in a lamp-holder fixed either in the runway itself or in a separate block of concrete, so that it does not protrude above the level of the runway. The bottom of the holder is a ring, resting on the concrete and secured to it by two plates. These plates can be loosened so as to allow the holder to be turned on its axis. Two screws passing through lugs on the holder also enable the latter to be tilted to some extent about a horizontal axis. If struck by a passing aircraft, the hard glass bulb of such a lamp will splinter without damaging the aircraft; on the other hand, experience has shown that the bulb, with a wall $1^1/_2$ mm thick, is strong enough to ensure that it will not be damaged by stones scattered low over the ground by the aircraft using the runway.

This light, being of a type intermediate between raised and flush lights, involves the problem common to all flush lights, that is, the protection of the fitting from water seeping through the ground; hence a rubber socket is fitted round the lamp cap to provide a water-tight seal when once the lamp has been inserted. Below the clamping nut screwed to the top of the lamp holder is a rubber ring, which when the nut is tightened, presses against the rubber socket and against the inner wall of the lamp holder The effectiveness of the seal is tested in an installation specially designed to test outdoor lighting fittings. Part of this installation is shown in *fig. 18*, from which it will be seen that in order to simulate practical conditions the lights on test are mounted in a tank which is periodically flooded with water.

In case the replacement of lamps broken by the wheels of aircraft is felt to be an objection, but cleaning the lights in the event of a snow-fall presents no difficulty, flush runway lights will be preferred. This system is particularly suitable when the lights, in order to obtain better height indication during the last phase of the landing, are to be arranged in two parallel rows relatively close together, say, about 150 ft apart, instead of along the edges of the paved runway, which may be as much as 200 ft wide. The probability of collision is

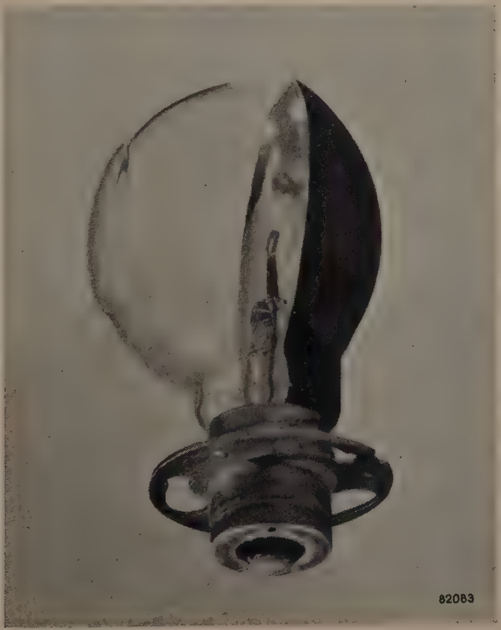


Fig. 19. Special 35-W lamp employed in a flush runway light. Part of the bulb, in the form of an ellipsoid of revolution, is silvered on the inside. Height of lamp: $2^1/_2$″.

then too great to permit of the use of elevated lights.

To ensure that flush lights will neither damage, nor be damaged by aircraft running over them, the mounting must be so designed that the lights do not protrude more than a few inches above the level of the runway. Philips have developed a special lamp enabling the required light beam to be obtained despite the above-mentioned limitation and without employing an expensive optical system (an important condition having regard to the fact that a large number of lights is required for each runway). This lamp is shown in *fig. 19*. One side (the back) of the bulb is an ellipsoid of revolution; the filament of the lamp is at one focus of this ellipsoid, and the image of the filament, whichs acts as the light source proper for the optical system, is at the other focus, outside the bulb. The optical system is a rectangular section cut from an aspherical lens of pressed glass (*fig. 20*). The "image" of the filament in the second focus of the ellipsoid really comprises a series of images projected one on top of the other by the different zones of the bulb-mirror and therefore differently magnified. This prevents any sharp definition of the filament structure either in the "image" or in the cross-section of the light beam, and furnishes a large image consistent with the required beam-spread.

Two systems of this type, fixed to adjustable brackets, are mounted back-to-back in the cover of a closed cast-iron housing, which is so embedded in the runway as to protrude less than 3″ above the latter; see *fig. 21*. This housing allows in a simple way the insertion of colour filters in the beam in order to code the beacon system. For further details see the captions of fig. 20 and fig. 21.

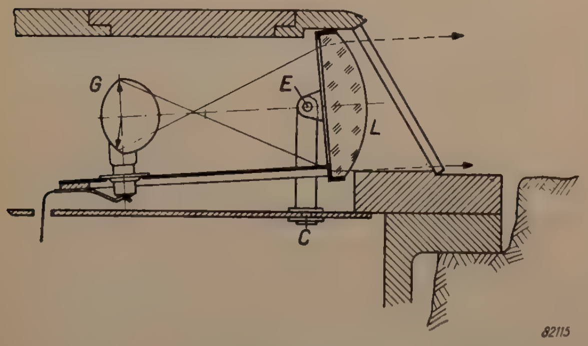


Fig. 20. Optical system of a flush runway light fitted with the lamp (*G*) shown in fig. 19. By virtue of the ellipsoidal mirror in the bulb, the light emitted by the filament is concentrated within a solid angle of nearly 2π, thus giving the required beam despite the fact that the height of the (square) lens (*L*) is only 55 mm. This runway light protrudes less than 3″ above the level of the "rough". To procure a beam of the same width and uniformity, with the same optical efficiency, by a more conventional optical system (e.g. that of the approach light, fig. 12), it would be necessary to employ a much larger system. The elevation of the beam is varied by tilting the optical system and the lamp on a horizontal spindle (*E*). Azimuth adjustment is by rotation about the vertical pin (*C*).

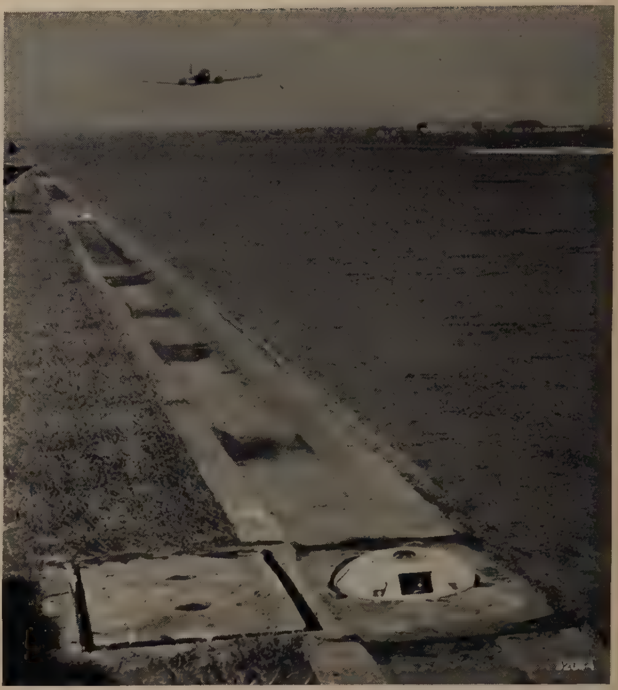


Fig. 21. Flush runway light on runway 05/23 at Schiphol. The light is housed in a cast-iron case, and contains two systems of the type shown in fig. 20, mounted back to back, and an omni-directional top light. To make the housing watertight, a rubber ring is inserted in a circular groove in the cover, close to the securing bolts. The lower part of the housing accomodates three supply transformers.

Threshold lights

When passing the threshold of a runway, aircraft in process of landing are still travelling at high speed; therefore, even the relatively slight elevation of 3 inches above the "rough" is not always considered permissible for lights situated at this point. Shallower lights of the type as that shown in fig. 21 have been designed for this purpose, but require a very special optical system to produce the required luminous intensity. The difficulty is enhanced by the fact that, according to the international recommendations, threshold lights should be green. To obtain green light with incandescent lamps it is necessary to employ a colour filter whose transmission can be only about 15%, so that in fact the luminous intensity of the beam before it passes through the filter must be in the region of 170,000 candelas.

Another solution to the problem, more attractive from the point of view of illumination engineering, is provided by what is known as a grid threshold light; see *fig. 22*. In this, the light supplying the required beam is below the surface of the concrete runway and protected by a steel coverplate. The light beam, projected upwards at a small angle, emerges through a slot cut in the runway in front of the light. Vertical plates of steel, strong enough to give adequate support to the wheel of

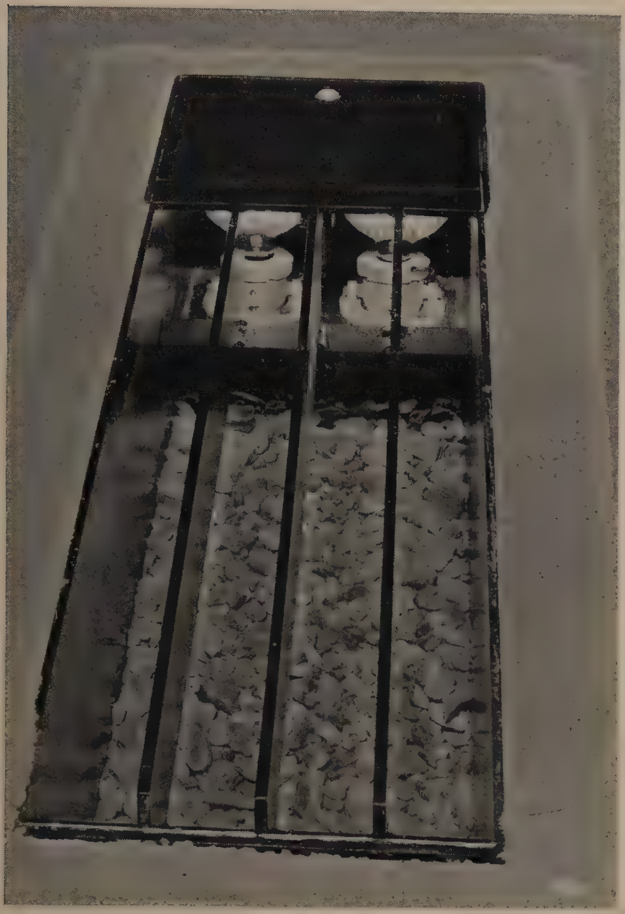


Fig. 22. Grid-type threshold light, fitted with two lamps of the type shown in fig. 15 and fig. 17.

any aircraft passing over them, are fixed in this slot. In the type of grid light illustrated, an internally silvered 100 watt lamp of the type shown in fig. 15 and fig. 17 is employed. Since it is possible to mount this lamp just below the cover-plate, the slot in the runway heed not be unduly long and the installation costs of the whole light can be kept within reasonable limits. Of course, the grid plates cause a certain loss of light. A spacing of these plates which is acceptable from the point of view of the width of aircraft wheels may well involve a light-loss of 20% along the axis of the beam, and of 50% in a direction deviating 5° from this axis. Hence each light should comprise several lamps arranged side-by-side to emit parallel beams, like the two lamps shown in fig. 22.

Aiming the beams

The required aiming of each individual light beam in a particular beacon system can be deduced direct from the diagram in fig. 10*a* ;the axis of the beam should pass through the centre of a cross-section of the visual guidance channel taken at a distance $r = 1500$ feet from the particular light (i.e. the required visual range) [7]. *Fig. 23* shows the required angle of elevation β, deduced from fig. 10*a*, plotted against the position of the light. The required azimuthal angle, α, depends entirely on the lateral distance b between the light and the centre-line of the runway ($\alpha = \tan^{-1} b/r$).

The means of azimuth and elevation adjustment provided in the different lights have already been referred to above. Two of the aiming devices designed to enable the lights to be adjusted quickly and conveniently to the required angles α and β will now be described.

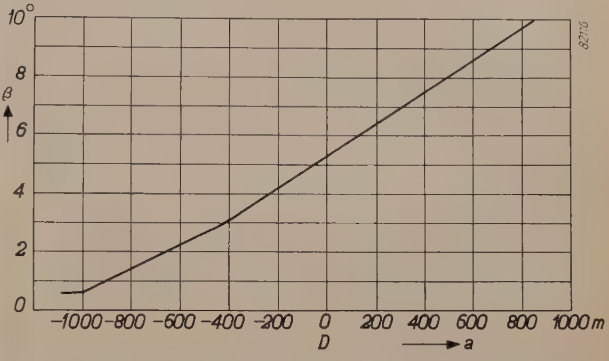


Fig. 23. The required angle of elevation (β) of the beam axis of individual lights in the approach and runway lighting system, plotted against the distance (a) between light and threshold.

Fig. 24 shows the device employed to aim the runway light illustrated in fig. 17. The device includes a prefocusing lamp-cap similar to that of the lamp to be fitted in this runway light. A moveable sight attached to the device can be adjusted to the required azimuthal angle α relative to the "beam-direction" of the prefocusing cap. The cap of the device is then inserted in the lampholder of the particular light, and this holder is rotated on its vertical axis until the sight is in line with the next light

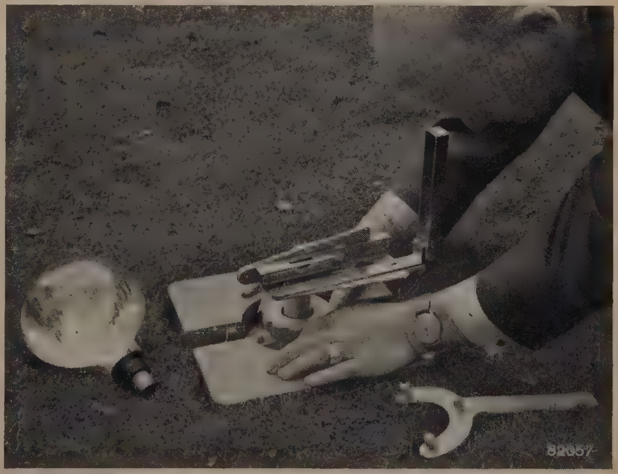


Fig. 24. Aiming device for runway lights of the type shown in fig. 17. The lamps used in such lights, being fitted with a prefocusing cap, will stand at an accurately pre-determined position when inserted in a lampholder. In view of this fact, the *lampholder* is aimed by means of this apparatus. The azimuthal deviation of the sight relative to the beam-direction the lamp will have can be varied. A prism is mounted in the sight to facilitate the adjustments by enabling the user to look down into the instrument.

---

[7] This principle is not always adhered to. In some systems of approach lighting, all the beam-axes are in vertical planes parallel to the centre-line of the runway, or even slightly toed-out. Such aiming necessitates greater beam-spread.

in the row (that is, parallel to the landing-direction); a holder so aligned will give the correct azimuthal angle to the beam of the particular lamp inserted in it. The instrument also includes a spirit-level, so mounted that it can be tilted along a graduated arc parallel to the "beam direction" of the lamp cap. If the spirit-level be tilted to the desired angle β and the lampholder then tilted until the spirit-level is again horizontal, the light will assume the correct angle of elevation.

The apparatus shown in *fig. 25* is employed to aim flush runway lights of the type shown in fig. 12. It consists of a large, square-cut lens, mounted on a frame, a peep-sight parallel to the centre-line of lens and frame, and a screen, calibrated in degrees of azimuth and elevation, at right angles to this centre-line. The lens is placed immediately in front of the beam-emitting aperture of the runway light; the frame is levelled with the aid of two spirit-levels, and its centre-line is lined-up parallel to the runway by sighting on the next light in the row or on a marker post placed at the edge of the runway. The bright spot of light produced by the lens as an image of the light-beam on the screen then indicates the azimuth and elevation of the beam. With the aid of the adjusting screws provided in the light, the optical system forming the beam is then rotated about horizontal and vertical axes until the light-spot strikes the correct point on the screen.

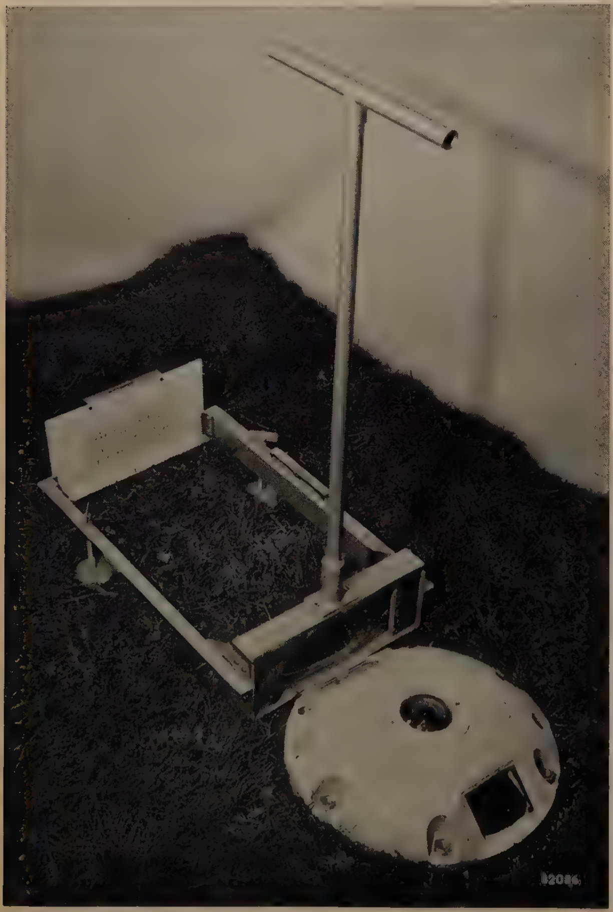


Fig. 25. Aiming apparatus for the flush runway light shown in fig. 21.

It may be worth mentioning, in conclusion, that the possibilities of further developments in landing techniques for aircraft are by no means exhausted. The present situation, that is, that an airfield must close down temporarily when the visibility drops below $V = 1000$ feet, will not be tolerated indefinitely. However, the solution to this problem lies less in increasing the luminous intensity of the beacons — as pointed out, this would be very expensive — than in the development of automatic control systems for the landing of aircraft.

Even when such a system is provided, the pilot will be loath to dispense with visual guidance entirely; however it will then be possible to reduce the altitude at present specified as the minimum for effective visual guidance. With the smaller visual guidance range then permissible, the present beacon system would be adequate in conditions of worse visibility. Indeed, it may probably be predicted that as far as the lights are concerned a point of development has now been reached where further major improvements are unlikely.

Summary. Apart from guidance by radar or radio beacons, which can home an aircraft to within about 3000 ft of the runway-threshold, the pilot coming in to land at night or in fog requires visual guidance to enable him to land safely. Such visual guidance is supplied by a system of beacons comprising approach lights, threshold lights (at the start of the paved runway) and runway lights. According to the recommendations of the I.A.T.A. and I.C.A.O., the approach lights should be arranged along the centre-line of the runway and along one or more transverse lines on this centre-line, and, given a visibility of only 1000 ft, all the lights should be visible, by night as well as in daytime fog, from a distance of at least 1500 ft in all directions within a certain "visual guidance channel". In accordance with these recommendations, it is calculated that the luminous intensity of all the lights should be at least 25,000 candelas, and the required light-distribution (isocandela diagram) of each light is determined. To satisfy these requirements, Philips have designed a series of special incandescent lamps and fittings. Lights belonging to each of the three classes referred to above are described in this article. The actual runway lights may be either "elevated" or "flush" the former type being so designed that if struck by an aircraft they will break without damage to the plane, whilst in the latter type special attention is given to water-tightness of the fitting. Threshold lights should preferably be entirely flush and should be green — a very stringent combination of requirements; one of the most suitable designs is the so-called grid-type threshold light. All the lights are provided with a means of varying the elevation and (where necessary) the azimuth of the beam: the beams of the lights installed on an airfield are aimed with the aid of special equipment.

FEEDBACK AMPLIFIERS FOR CARRIER TELEPHONE SYSTEMS

by J. te WINKEL.

621.375.232:621.395.44

In carrier telephony systems used for long-distance communications, techniques have been developed whish permit a large number of speech channels — up to some hundreds — to be carried on each line. The very wide frequency bands which then have to be transmitted by the cable introduce some interesting problems regarding the design of the line amplifiers that have to be included at regular intervals in the cable circuit.

Introduction

To cope with the vastly increased long-distance telephone traffic it is expedient for economic reasons to transmit a number of channels simultaneously over one pair of conductors. For this purpose each audio signal, usually restricted to frequencies between 300 and 3400 c/s, is modulated onto a different H.F carrier frequency, the result being effectively a displacement of the frequency band. The number of channels that can thus be transmitted over one conductor pair when these bands are placed side by side in the frequency spectrum is determined by the total bandwith that can be transmitted by the circuit. In modern telephone cables, containing a number of conductor pairs and termed symmetrical cables (as distinguished from the coaxial cables to be mentioned later) the number of channels operated over each conductor pair is nowadays as much as 48 or 60. The limit to this number of channels, and hence to the width of the total frequency band to be transmitted, is entirely determined by the properties of the cable. Crosstalk is fundamentally the limiting factor, since this increases with increasing frequency and thus also when the number of channels is increased [1]). Special means, which will be discussed later, are always used to compensate this crosstalk as much as possible. The remaining amount, however, still constitutes an upper limit to the frequency. The designer of amplifiers for use in conjunction with this type of cable should therefore consider the total bandwidth as the primary design parameter. If it is desired to carry even more channels over one pair of conductors by means of conventional single-sideband modulation, other types of cable have to be employed. The coaxial cable is particularly suitable for this purpose. In this type of cable, the inner conductor is completely surrounded by the other, the tubular outer conductor. Owing to the skin effect in the outer conductor hardly any external field is created (provided the frequency is not too low), so that no crosstalk occurs. In this respect there is no limit to the number of channels and thus to the total bandwidth. Since, however, the attenuation increases with the frequency also in this type of cable, the degree of amplification must become greater the wider the frequency band, or the amplifiers have to be spaced at shorter intervals. This constitutes a fundamental limitation on the possible number of channels. In practice bands up to 4 to 8 Mc/s wide, with room for 1000 to 1800 channels, can be used. This large number of available channels is in itself an advantage of the coaxial cable. On the other hand it gives rise to several particular difficulties, so that in practice both coaxial and symmetrical cables have found their own fields of application [2]). The amplifiers which have to be incorporated in the transmission line at regular intervals, known as line amplifiers, are the subject of this article. They have to meet a number of specific requirements. The nature of these requirements, and under what conditions they can be satisfied will be explained below. Finally certain types will be examined more closely.

Design criteria for line amplifiers

The various properties to be considered when designing a line amplifier for carrier telephony are as follows:

1) Large absolute bandwith
2) Large relative bandwidth
3) Very constant amplification (high stability).

---

[1]) See for example, G. H. Bast, D. Goedhart and J. F. Schouten, A 48-channel carrier telephone system, Philips tech. Rev. 9, 161, 1947/48.

[2]) For a more detailed discussion of the two systems, see H. N. Hansen and H. Feiner, Coaxial cable as a transmission medium for carrier telephony, Philips tech. Rev. 14, 141-150, 1952/53.

4) Amplification to be independent of the amplitude of the signal (linearity).
5) Sufficient degree of power amplification.
6) Well-defined input and output impedances (i.e. the impedances presented by the amplifier to the line).

These requirements will now be discussed in turn.

Large absolute bandwidth

The absolute bandwidth, i.e. the difference between the highest and the lowest frequency of the signal to be amplified, is larger according to the number of channels used. With symmetrical cables, for the transmission of 48 channels, the band 12-204 kc/s is used, and for 60 channels the band 12-252 kc/s With coaxial cables the permissible bandwidth is far greater and is, as mentioned in the introduction, limited only by the properties of the amplifier. At the present state of amplifier technique, it is possible to design amplifiers for the bands up to 60-4100 kc/s or even 0.3-8 Mc/s.

These bandwidths are of the same order of magnitude as those required for the transmission of a TV signal (5 Mc/s for the 625-line standard). The above-mentioned amplifier for 0.3 to 8 Mc/s has been developed also with a view to this application. It has, therefore, to meet certain additional requirements applicable to TV transmission, but these will not be considered here.

Large relative bandwidth

The relative bandwidth, i.e. the ratio of the highest to the lowest frequency, will always have to be large, because at a given absolute bandwidth it will always be attempted to keep the lowest frequency (and hence also the highest) as low as possible in view of the fact that the attenuation increases with the frequency for both types of cable. For the above-mentioned bands of 0.3-8 Mc/s and 60-4100 kc/s the relative bandwidth amounts to approx. 27:1 and 70:1 respectively.

A large relative bandwidth causes the following difficulties: The amplifier includes a number of components, such as inductances and capacitors, whose values and hence whose dimensions have to be greater as the lower frequency limit becomes lower. The larger their dimensions, the more are all these components liable to give rise to all sorts of secondary effects (e.g. the self-capacitance of a coil, the leakage inductance of a transformer, the earth capacitance of a capacitor, etc.). As a rule these secondary effects either directly or indirectly reduce the amplification at higher frequencies. Clearly this constitutes a limitation on the relative bandwidth. The designer of components can make a useful contribution here by reducing the disturbing effects, for example by using new materials and thus decreasing the dimensions.

High stability

Several factors may cause the amplification to vary. Fluctuation of temperature and supply voltages may influence the properties of several components. This may also occur without external influences, simply by a process of ageing, particularly where valves are concerned.

As regards the amplification necessary in coaxial circuits, strict limitations are imposed upon amplification variations. Now, it has been found in practice that the distance between two repeater stations must never exceed 10 km. If it were possible to make each amplifier constant within 1% (which means an amplification variation of ± 0.09 dB), then a succession of 100 amplifiers (a transmission distance of 1000 km is by no means exceptional) would result in a variation that could amount to as much as ± 9 dB. It is true that since the attenuation of the cable varies with the temperature, it is anyhow necessary to provide for a certain compensation — either automatic or not — for the fluctuations in amplification of the whole system of cable and amplifiers. In view of the fact that this compensation introduces complications, however, it is advisable not to compensate more than is strictly necessary, so that fluctuations in amplification should be kept as small as possible.

For amplifiers in transmission systems using symmetrical cables this requirement is less strict, because here the repeaters may be spaced at intervals of 18-25 km, so that a smaller number is required to bridge a given distance.

Amplification independent of the amplitude (linearity)

If the instantaneous output voltage of an amplifier is not a purely linear function of the instantaneous input voltage, this gives rise, in the case of a complex input signal, to harmonics and sum-and-difference frequencies in the output signal. If many channels are amplified at the same time, interference signals will appear roughly uniformly distributed throughout the band, which may be regarded as noise. This is called *intermodulation noise* because it is mainly caused by the interaction of different frequencies. Although the sum and difference frequencies are as a rule of greater importance, it is possible to calculate the intermodulation noise from measurements of the second or higher harmonics using a single variable frequency [3]). From this the quantitative requirements regarding the linearity of the amplifier can be derived.

---

[3]) See R. A. Brockbank and C. A. Wass, J. Inst. El. Engrs. III, **92**, 45-56, 1945.

The intermodulation noise can be reduced by keeping the amplitude of the input signals very small, but this remedy is restricted by the fact that noise is also caused by other factors (thermal noise, valve noise) already present at the amplifier input. This noise is independent of the signal level. It thus sets a certain lower limit to the input voltage if one is to maintain a certain miminum signal-to-noise ratio.

Power amplification

The production of sufficient power does not constitute a problem in itself. There are several types of amplifying valves, up to the large transmitting valves, that are capable of dissipating large powers. Raising the output power, however, requires larger valves as well as larger output transformers and thus gives rise to the disturbing secondary effects mentioned under (2). In practice, moreover, it has been found that there is little point in raising the output power above a certain level, as we shall now demonstrate.

The purpose of the amplifier is obviously to compensate the attenuation occurring in a certain length of line. As an example let us consider a coaxial cable with line amplifiers spaced at intervals of 10 km. Each amplifier has a gain of 50 dB and a maximum output power of 500 mW. If the voltage amplification is to be increased by a factor of $\sqrt{10}$, i.e. 10 dB, then, in view of the fact that the input signal is practically fixed, the output power must become 10 times greater, i.e. 5 W.

This would make the amplifier considerably more complicated and also require a larger power-supply. If all these objections were acceptable, then we would have obtained an amplification of 60 dB. The interval between the repeaters could then be enlarged to

$$\frac{60}{50} \times 10 = 12 \text{ km},$$

which does not make very much difference in the number of amplifiers required. To obtain 20 dB more amplification and a spacing interval of 14 km, an output power of 50 W would be necessary.

Input and output impedance

With amplifiers for symmetrical cables the impedance, presented to the cable by the input and output terminals, must be equal to the characteristic impedance of the cable. This is necessary in order to keep crosstalk at a low level, as may be seen with the aid of *fig. 1*. Lines *1* and *2* represent two adjacent conductor pairs. At the ends marked *B* are the output terminals of two line amplifiers; at *C* are the input terminals of the two next amplifiers. Let us assume that at *A* a crosstalk signal from conductor pair *2* is transferred to conductor pair *1*. This signal will be propagated in both directions. The part going to the right will arrive at *C* together with the original signal on conductor pair *2* irrespective of the position *A* where the signal originated.

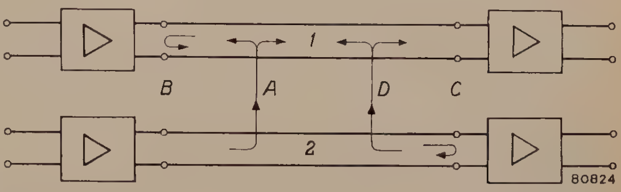


Fig. 1. Two conductor pairs (*1* and *2*) of a symmetrical cable link. At *B* are the output terminals of two amplifiers, at *C* the input terminals of the next two amplifiers. A signal producing crosstalk from *2* to *1* at *A*, will be propagated in *1* in both directions. That going to the right (together with the initial signal) can be compensated at *C*, because it is in phase with the initial signal. The part going to the left must not be reflected at *B*, for the phase difference (which depends on the position of *A*) cannot be compensated. Nor must reflection occur at *C*, since the crosstalk from this reflected signal (e.g. at *D*) cannot be compensated either.

By using additional adjustable coupling elements between the two conductor pairs this crosstalk signal can be fairly well compensated for all positions of *A* (this is known as *balancing* of the conductor pairs). The part of the crosstalk signal going from *A* to the left and being reflected at *B*, however, cannot be compensated at *C*, because it has become out of phase with respect to the direct signal, due to its longer path. In order to keep the crosstalk level as low as possible these reflections should be avoided whenever possible by terminating the conductor pairs at *B* with their characteristic impedances (matching). The termination at *C* should be similarly matched to avoid reflection of the signal itself. This reflected signal can also give rise to crosstalk, as indicated by *D* in the figure. This crosstalk signal also cannot be compensated.

In the case of amplifiers for coaxial cables, where crosstalk becomes negligible at higher frequencies, such matching is, as a rule, not required for telephony transmission (although it is necessary for TV signals, but for different reasons).

Design of the amplifier; negative feedback

One is thus faced with the following problem: a reliable, constant and linear amplification has to be effected over a wide band. A solution for this problem, well-known in amplifying technique, is the application of considerable negative feedback. The subject of negative feedback has been discussed more than once in this Review [4]), so it will suffice here to recall briefly its principle. *Fig. 2* shows an amplifier A and a feedback circuit B. The gain factor of A is μ. Part of the output voltage V_u is tapped off to the circuit B, in such a way that a part βV_u is deducted from the input signal V_i. The following relationships then exist between the effective input voltage V_i' and the output voltage V_u of the amplifier A:

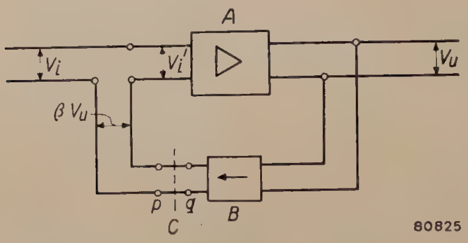


Fig. 2. Amplifier A with feedback circuit B. A fraction βV_u of the output voltage V_u is tapped off by the circuit B and applied in series with the input voltage of A (in opposite phase). The total amplification thus becomes more linear and less dependent on variable quantities.

$$V_i' = V_i - \beta V_u,$$
$$V_u = \mu V_i'.$$

and hence:

$$V_u = V_i \cdot \frac{1}{\beta} \cdot \frac{\mu\beta}{1+\mu\beta}.$$

If $\mu\beta$ is large compared to 1, the voltage gain becomes approximately equal to $1/\beta$ and, therefore, no longer dependent on μ, i.e. on the valves and other fluctuating or distortion-causing elements in the separate stages of the so called μ-circuit. β on the other hand, is determined by the ratio between impedances or between the numbers of turns on a transformer, and can thus readily be made constant and linear. A more rigorous treatment shows that all variations in amplification caused by fluctuations in valves and other components of the μ-circuit, as well as the non-linearity, are reduced by a factor $1 + \mu\beta$.

The relative change of amplification caused by a change $\Delta\mu$ is, without feedback, $\Delta\mu/\mu$, and with feedback:

$$\frac{\Delta\mu \cdot \frac{\mathrm{d}}{\mathrm{d}\mu}\left(\frac{\mu}{1+\mu\beta}\right)}{\frac{\mu}{1+\mu\beta}},$$

which is

$$\frac{1}{1+\mu\beta} \cdot \frac{\Delta\mu}{\mu}.$$

The change $\Delta\mu$ may have been caused by variations of any component in the amplifier μ-circuit.

It will thus be our aim to make the product $\mu\beta$, termed the feedback factor, as large as possible. The resulting amplification, i.e. the gain between input and output terminals, which, as stated, is nearly equal to $1/\beta$, is determined by the design of the whole transmission system. Hence μ should be given the highest possible value. We find, however, that the large bandwidth imposes certain limitations upon μ, which will be discussed below.

Limitation of the amplification at large bandwidth

For a single amplifying stage, which in simplest form may be considered to be a valve with slope S and anode impedance Z, the amplification amounts to SZ (so long as Z is small compared to the internal resistance of the valve).

One component of Z is due to the total stray capacitance $C_p = C_1 + C_2 + C_3$ (*fig. 3*), i.e. the sum of the anode-cathode and grid-cathode capacitances and the wiring capacitance. It is not possible to connect anything in series with this capacitance, but a network of impedances may be arranged in parallel with it. This impedance should be chosen so as to obtain the greatest possible amplification over the

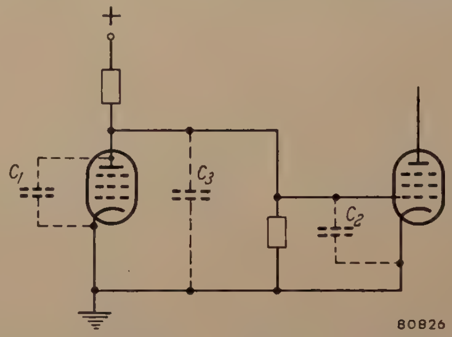


Fig. 3. Part of an amplifier. For determining the anode impedance, the following stray capacitances have to be taken into account:
C_1: the anode-cathode capacitance of the valve,
C_2: the grid-cathode capacitance of the next valve,
C_3: the wiring capacitance.

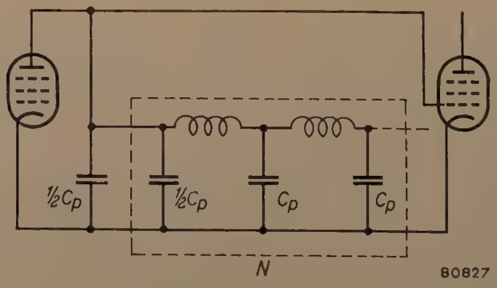


Fig. 4. Circuit for obtaining the optimum amplification for a given bandwidth. N is a low-pass filter with a cut-off at the highest frequency to be amplified. The first element of N is formed by half the stray capacitance $C_p/2$.

---

[4]) See, for example, C. J. van Loon, Philips tech. Rev. 1, 264-279, 1936; B. D. H. Tellegen, Philips tech. Rev. 2, 289-294, 1937.

given bandwith. According to Bode [5]) this should be done as follows: The stray capacitance is imagined to be divided into two equal parts, one part forming the capacitance of the first section of an ideal low-pass filter N (see *fig. 4*), the cut-off frequency of which is to correspond to the highest frequency f_m to be transmitted. It is known from the theory of filters that the admittance Y_p of such a filter at a frequency f is given by the general formula:

$$Y_p = Y_0 \sqrt{1 - \frac{f^2}{f^2_m}}, \quad . \; . \; . \; . \; (1)$$

where Y_0 represents the D.C. admittance.

For the case of this filter $Y_0 = \pi f_m C_p$.

The total admittance Y in the anode circuit is therefore:

$$Y = Y_p + j\pi f C_p = \pi f_m C_p \sqrt{1 - \frac{f^2}{f^2_m}} + j\pi f C_p. \; (2)$$

For $f \leqq f_m$, the modulus of this is:

$$|Y| = \pi f_m C_p.$$

The amplification μ then becomes:

$$\mu = S|Z| = \frac{S}{|Y|} = \frac{S}{\pi f_m C_p},$$

and is thus constant for all frequencies between 0 and f_m. We see that the absolute values of the amplification and of the impedance become uniform throughout the whole range. This impedance is equal to double the impedance of the stray capacitance at the cut-off frequency. This can be readily understood if we consider that the impedance of the filter is infinitely high at the cut-off frequency (cf. equation 1). Half the stray capacitance has, as it were, been removed from the circuit.

An ideal filter contains an infinite number of elements: in practice this has to be approximated to by a distributed network of a finite number of elements. With such a distributed network it is possible to attain 80-90% of the theoretical maximum amplification. When this is taken into account, it is found that the amplification, according to the above considerations, is determined by the highest frequency f_m and by the ratio S/C_p, and the latter cannot possibly exceed $S/(C_1+C_2)$ (see fig. 3). The ratio $S/(C_1+C_2)$, which is also an important quantity in other amplifying circuits, is termed the figure of merit of the valve in question. For the sake of simplicity this will be denoted henceforth S/C.

Once we have selected a tube with the best possible figure of merit, the maximum attainable amplification per stage is more or less settled. We should then employ as many stages in series as are necessary for the total amplification μ to reach the required high value. If more than two stages are necessary, which is very probable where coaxial amplifiers are concerned, certain complications may arise, due to the fact that the amplifier begins to oscillate spontaneously. This effect can be explained as follows. An amplifier with negative feedback possesses a second amplification circuit, the $\mu\beta$-circuit, which we can imagine if the feedback circuit is interrupted, say, at C in fig.2, thus creating the terminal pairs p and q. It is assumed that at p a slight potential difference v occurs, caused e.g. by the stray pick up of an interfering signal, whether the input voltage V_i at that particular moment is zero or not. Owing to the feedback, a voltage $\mu\beta v$ then prevails at q. This circuit has the inherent property (on which the whole idea of negative feedback is based) that the voltage at q is in anti-phase with the voltage at p. It is clear, therefore, that if at some frequency the output voltage is given an additional phase shift of 180° and thus be comes in-phase with the input voltage the whole system will start oscillating.

The precise conditions under which oscillation will occur are given by Nyquist's criterion [6]), which we shall recall briefly here. The ratio of the output to the input voltage in the feedback circuit of an amplifier is, as a rule, a complex quantity, which can be plotted as a point in the complex plane. This quantity is a function of the frequency, and as the frequency is varied the point in the complex plane describes a curve, generally a closed curve, since the amplification both at very high and at very low frequencies approaches zero.

Nyquist showed that a negative feedback amplifier will not oscillate unless this curve encircles the point -1 on the real axis [7]). In practice a more stringent criterion must be maintained, which is in effect the condition stated above, viz. the phase shift of the output must be less than 180° when the amplification is greater than unity.

Considering the phase shift between input and output voltage per stage, the risk of oscillation is considerable. At low frequencies the anode impedance will behave very nearly like a resistance. At frequencies equal to or higher than f_m the maximum frequency to be transmitted, equation (2) shows that the anode impedance is purely imaginary, so that at frequencies equal to or higher than f_m the phase shift is 90° more than at low frequencies. Using two valves, we would thus just reach the condition for oscillation viz. 180° extra phase shift in the $\mu\beta$-circuit. With more than two valves the phase difference might be even greater and, with an amplification in the circuit which could be greater than unity, the amplifier would promptly start oscillating.

---

[5]) H. W. Bode, Network analysis and feedback amplifier design, Van Nostrand, New York, 1945.

[6]) H. Nyquist, Bell Syst. tech. J. **11**, 125-147, 1932.

[7]) See also the article by B. D. H. Tellegen referred to in [4]).

Here we encounter a new problem: how to obtain in this second amplifier circuit — i.e. from p to q in fig. 2 — not only the greatest possible amplification $\mu\beta$ over a given frequency range, but at the same time a phase shift which does not exceed 180° for all those frequencies for which $\mu\beta$ is greater than unity. This problem too has been studied theoretically by Bode. His study shows that for each network, irrespective of its circuitry or number of valves there must exist a certain relationship between the amplification and the phase shift. To a first approximation the phase shift is proportional to the ratio of the change of the logarithm of the amplification to that of the logarithm of the frequency, i.e.

$$\varphi = c \frac{\mathrm{d} \log \mu\beta}{\mathrm{d} \log f} . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The exact formula may be found in the book referred to in [5]): the above formula (3) applies if $\mathrm{d} \log \mu\beta / \mathrm{d} \log f$, i.e. φ, varies relatively slowly with the frequency.

In order to avoid oscillation of the amplifier, φ should be less than 180° for all values of $\mu\beta$ greater than unity. To give a concrete example, we assume φ to be equal to a constant φ_0 for all frequencies between f_{m}, the highest frequency transmitted, at which $\mu\beta$ has nearly the maximum value, and the frequency f_0, at which $\mu\beta$ is unity. We can then integrate equation (3) between f_0 and f_{m} and find:

$$\mu\beta_{\max} = \left(\frac{f_0}{f_{\mathrm{m}}}\right)^{-\frac{\varphi_0}{c}} .$$

An exact calculation shows that for $\varphi_0 = -180°$ and for three stages of amplification,

$$\mu\beta_{\max} = \left(\frac{4}{3} \frac{f_0}{f_{\mathrm{m}}}\right)^2 .$$

The parameter f_{m} may be considered as given. The maximum value of $\mu\beta$, i.e. the maximum feedback factor between $f = 0$ and $f = f_{\mathrm{m}}$, is thus known if we know the frequency f_0. This frequency, for which $\mu\beta$ has dropped as low as unity, is so high that the anode impedances of the various stages are nearly equal to the stray capacitances.

As a concrete example, let us assume that for this frequency $\beta = 1$, then also $\mu = 1$ and the amplification per stage is 1, so that:

$$\frac{S}{2\pi f_0 C_{\mathrm{p}}} = 1, \text{ or } f_0 = \frac{1}{2\pi} \frac{S}{C_{\mathrm{p}}} .$$

Since the value of β is generally somewhat smaller than 1, dependent on the circuit, f_0 will assume a slightly lower value, but the essential fact remains that f_0 bears a direct relationship to S/C ,the figure of merit of the type of tube in question.

We now arrive at the following result. For a given type of valve with a given S/C value, the maximum attainable feedback $\mu\beta$ is determined by the highest frequency to be transmitted f_{m}, $\mu\beta$ decreasing as f_{m} increases.

We see that enlarging the bandwidth is achieved at the expense of the feedback and hence at the expense of the linearity and of the stability of amplification. The better the quality of the valves, expressed as the figure of merit, the larger bandwidth and/or feedback is permissible.

Having dealt with the fundamentals of these amplifiers we shall now consider some practical designs.

Practical designs

Amplifier for symmetrical cables (12-204 kc/s)

In the foregoing we have shown that the order of magnitude of both input and output signal is more or less determined. The input level has to lie a certain amount above the total noise level and the output power should be such that it can be delivered by a tube of normal dimensions. The magnitude of the resulting amplification $1/\beta$ is therefore approximately determined. If, furthermore, the feedback factor $\mu\beta$ is specified, μ can be found. μ is the amplification without feedback and hence the product of all amplifications per stage. Its numerical value is such that two amplifying stages, having equal amplification for all frequencies of the band, are not sufficient. This requirement of a uniform amplification for all frequencies, however, is not essential; the fact that higher frequencies are attenuated to a greater extent by the cable than the lower ones makes it permissible for $1/\beta$ to decrease at lower frequencies. If the feedback is to be kept constant, μ also has to decrease in proportion to the frequency, in fact, in a ratio that is inversely proportional to β This can be realized by means of complementary networks in the μ and β circuits [8]).

Owing to the fact that μ decreases with decreasing frequency, it is possible to reach higher values of μ at higher frequencies than could be achieved if μ were uniform throughout the frequency band, and in this way two amplifying stages will be sufficient. This means a considerable saving as compared to a 3-stage amplifier. In addition, the requirement of stability can be better satisfied, as demonstrated above.

---

[8]) See H. van de Weg, The equalization of telephone cables, Philips tech. Rev. 7, 184-191, 1942.

Fig. 5 shows the rising frequency characteristic within the band 12-204 kc/s. At 204 kc/s the amplification is 68 dB. The feedback factor $\mu\beta$ is 32. With present symmetrical cables and a band of 12-204 kc/s these repeaters can be placed at intervals of 25 km.

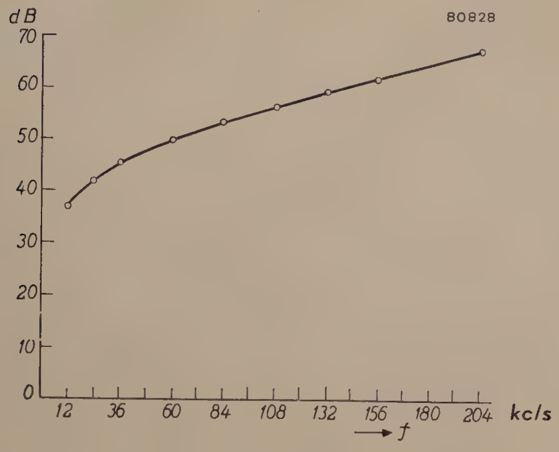


Fig. 5 The amplification of the amplifier for symmetrical cables as a function of frequency. In order to use only two amplifying stages, the amplification is made to rise with the frequency.

The amplifier under consideration uses a combination of current and voltage feedback in order to match its output impedance to that of the cable and thus to keep the cross-talk level low. Due to the voltage feedback (as shown in fig. 2) the terminal voltage will be kept nearly constant by the amplifier uring variations of the load resistance. The internal resistance measured at the output terminals is low in this case. In the case of current feedback a voltage is returned to the input terminals that is proportional to the output current. As a result of this, variations in the output current are suppressed, so that the internal resistance is high. Without any calculations, it will be seen that by combining current and voltage feedback any desired resistance may be obtained. (Matching could also be effected with only voltage feedback or only current feedback by introducing an additional parallel or series resistor, but this would always result in a loss of output power.)

Fig. 6 shows a simplified diagram of the amplifier with the combined feedback. Current feedback is established across resistor *R* and voltage feedback through the secondary winding *S* of the output

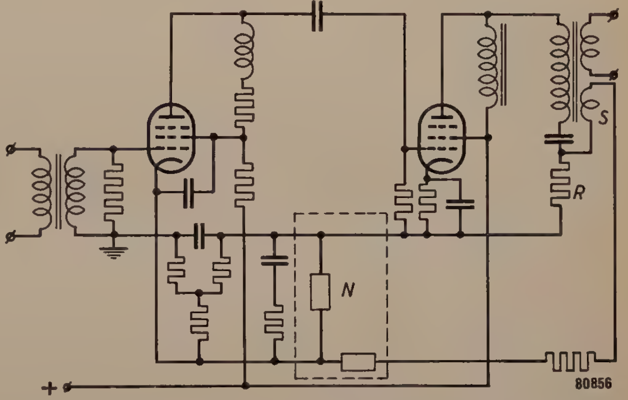


Fig. 6. Simplified diagram of a two-stage amplifier for symmetrical cables. Current feedback is provided by resistor *R*, whilst voltage feedback is effected by the secondary winding *S*. The equalizing network *N* ensures a straight frequency-response curve of cable plus amplifier.

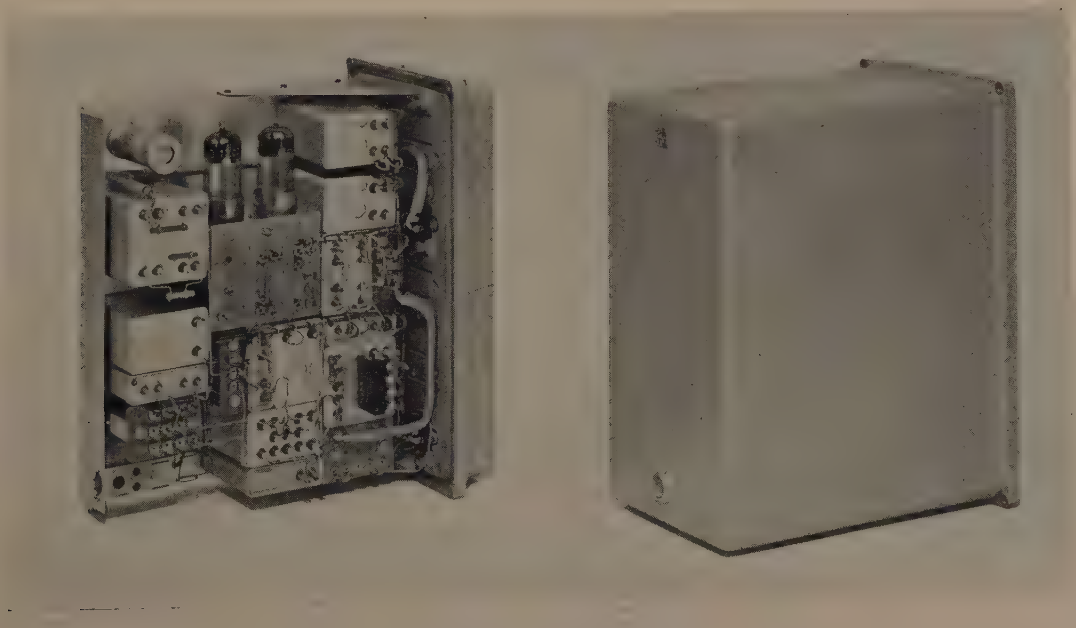


Fig. 7. The two-stage amplifier for symmetrical cables.

transformer. The equalizing network *N* has the function of straightening out the overall response curve of cable plus amplifier; it varies $1/\beta$ in such a way that at all frequencies the resulting amplification just compensates the cable attenuation. *Fig. 7* is a photograph of this amplifier.

Amplifier for coaxial cables (300-4200 kc/s)

As a second example we will consider a line amplifier for coaxial cables. This amplifier has been designed for the frequency band 300-4200 kc/s and is built up of three stages. The highest frequency is so much higher than that in the previous example that even by means of an amplification increasing with frequency the value of μ that could be reached with a two-stage amplfier would be too low. The diagram of *fig. 8* shows that only current feedback across the impedance *Z* has been applied. It should further be noticed that the coupling elements between the tubes are rather complex. This is necessary in order to obtain the closest possible approximation to the aforementioned theoretical value of the feedback factor $\mu\beta$.

All valves have a second identical valve in parallel, a necessary measure in view of the fact that in this case a single amplifier has to deal with 900 channels. *Fig. 9* shows what happens if one, two or three of the additional parallel valves are put out of action for some reason. Even in the extreme case of three non-effective valves, the change in amplification will not exceed approximately 1 dB, thanks to the feedback. We see that the overall amplification between 300 and 4200 kc/s is approximately 41 dB and if all valves are functioning, remains constant

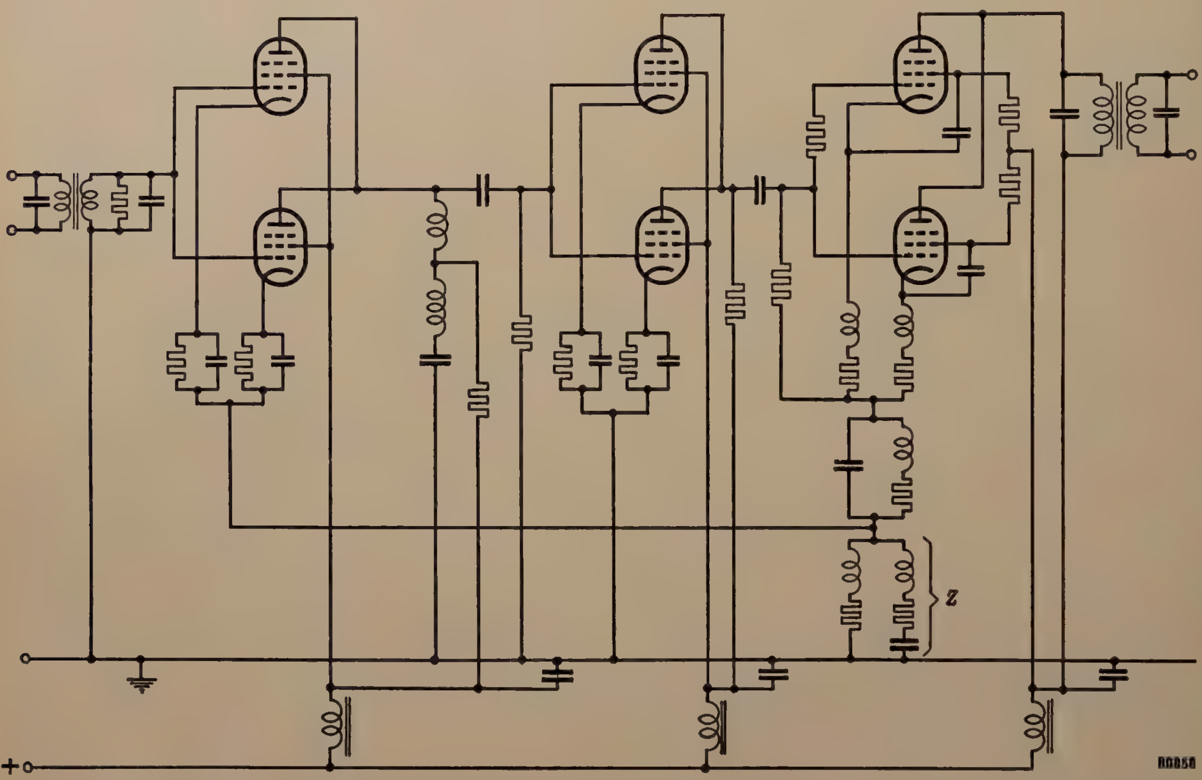


Fig. 8. Simplified diagram of the amplifier for coaxial cables. Current feedback is obtained across the impedance *Z*. All tubes work as parallel pairs. The coupling elements are such that approximately 80% of the theoretical maximum value of $\mu\beta$ can be attained.

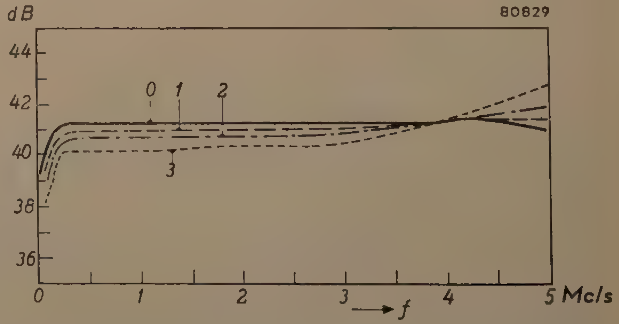


Fig. 9. Amplification as a function of the frequency for the amplifier for coaxial cables: *0*, all valves functioning (amplification is then constant within 0.1 dB throughout the range 0.3-4.2 Mc/s); *1*, one valve in one stage failed; *2*, one valve in each of two stages failed; *3*, one valve in each of three stages failed. Even in the last case the deviation is no more than 1 dB.

within 0.1 dB. We have demonstrated above that both the amplitude and the phase of the feedback factor $\mu\beta$ are of importance. This factor has therefore been plotted in a polar diagram in *fig. 10*. We see that in the band to be amplified $\mu\beta$ is almost constant and amounts to 30 dB (31×), but that at higher frequencies, at a phase angle of about −150° the amplitude decreases to below 1, so that the point (−1, 0) (amplification 1 at phase angle 180°) falls entirely outside the curve. Thus the condition for non-oscillation is satisfied. We see that in this respect also the behaviour of the amplifier at frequencies up to approximately 50 Mc/s is important [9]). At these frequencies parasitic effects may be very troublesome. A wire of 10 cm length then represents an inductive impedance of about 30 Ω. Very careful assembly is thus necessary and, as a rule, all connections should be kept as short as possible. The photographs of this amplifier in *fig. 11* give an impression of the assembly arrangement. The valve-holders are as near to each other as possible and the various coupling elements are mounted between them. Capacitative coupling between the valves has been avoided by building into the chassis two mounting plates, in the form of a cross, and by locating the valve holders in a suitable way with respect to these two crossed plates.

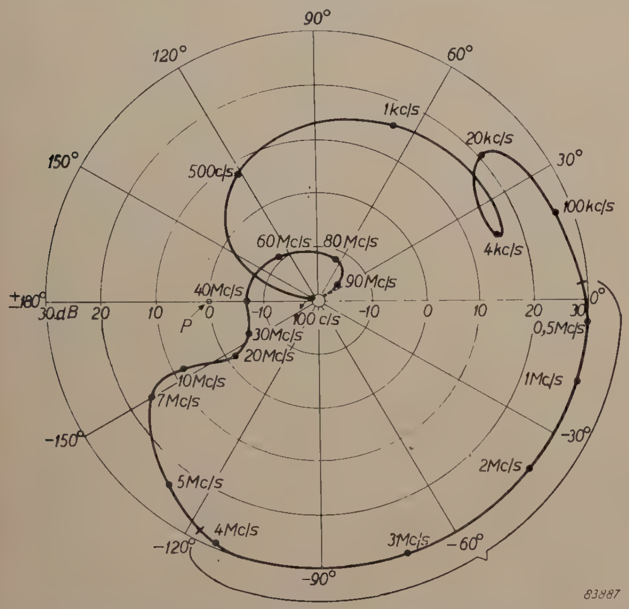


Fig. 10. Polar plot of the value $\mu\beta$ of the amplifier for coaxial cables. In the band to be amplified $\mu\beta$ is approximately 31; this range is indicated by the bracket. In view of the rather large variations to be shown, the amplitude has been plotted logarithmically (in dB) (20 dB corresponds to a factor of 10). Point P (0 dB, 180°) is clearly outside the curve, so that the amplifier, according to Nyquist's criterion, cannot oscillate. It can be seen that in order to be sure of this, it is necessary to know the value of $\mu\beta$ for frequencies up to about 50 Mc/s.

Summary. In carrier line-telephony attenuation in the cable makes it necessary to place line amplifiers at regular intervals along the cable, capable of highly stable and linear amplification over a large bandwidth. It is found that, at a given bandwidth, both the coupling impedance between the valves of such an amplifier and the amount of the amplification per stage are mainly determined by the figure of merit S/C of the valves. In order to improve the linearity and to reduce amplifi-

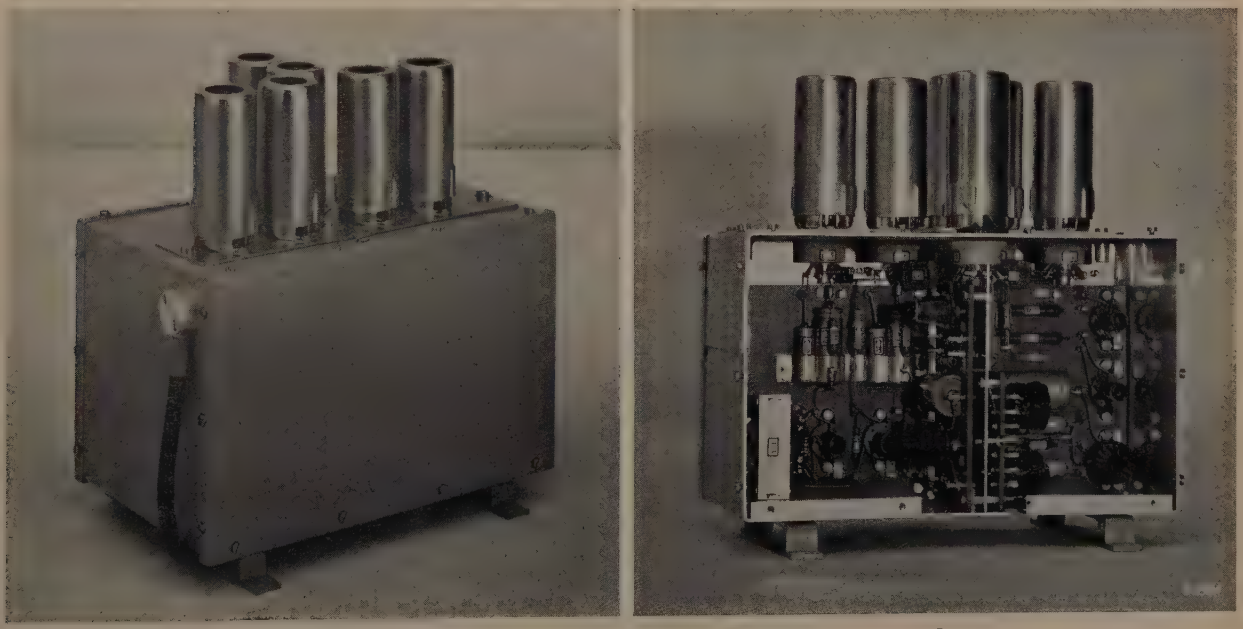

a
b

Fig. 11. Amplifier for coaxial cables. *a*) Front. *b*) Rear.

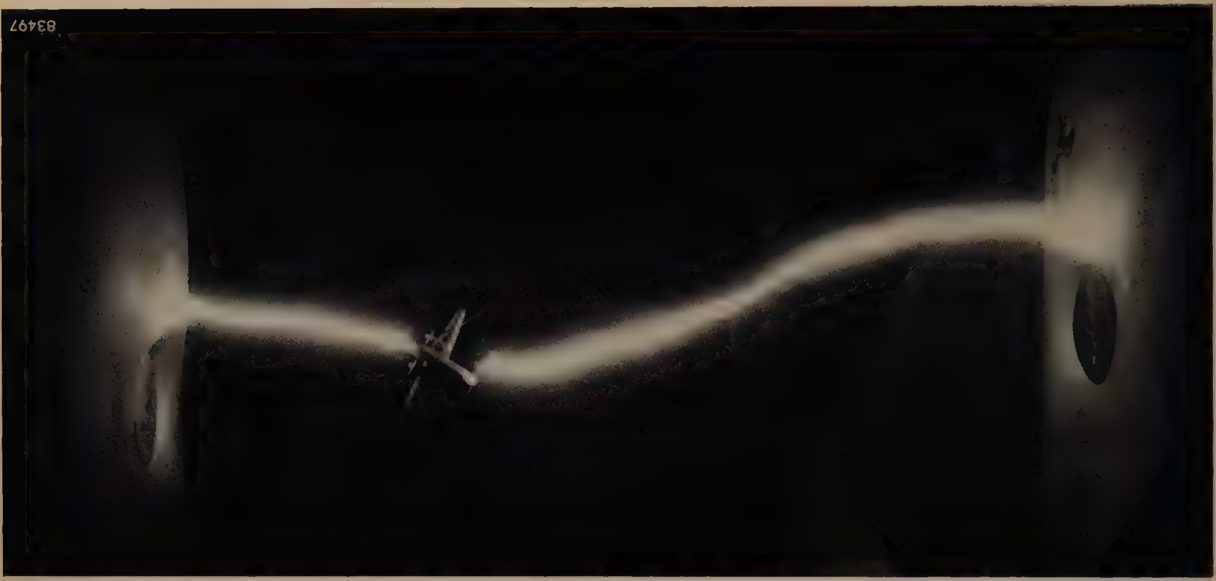
---

[9]) See also H. Thirup, An instrument for measuring complex voltage ratios in the frequency range 1-100 Mc/s, Philips tech. Rev. **14**, 102-114, 1952/53.

cation variations, feedback is applied. The article demonstrates that S/C similarly limits the amount of the feedback, which, if exceeded, may put the feedback in anti-phase at very high frequences, so that the amplifier oscillates. These considerations are applied to two types of cable currently used in modern carrier telephony — the symmetrical and the coaxial cable. An amplifier for each of these types is described. The amplifier for symmetrical cables operates in the range 12-204 kc/s, which is the bandwidth required for 48 channels. By using an amplifier with a rising frequency characteristic, it is possible to use only two amplifying stages, A correct output impedance is achieved by a combined current and voltage feedback. The maximum amplification is 68 dB at a feedback factor of 32. The amplifier for coaxial cables operates in the band 0.3-4.2 Mc/s (which can accomodate a good 900 channels). It has a constant amplification of 41 dB at a feedback factor of 31. Three stages are necessary in this amplifier. To ensure a high degree of reliability all valves are paired. In the construction of this amplifier special steps are taken to reduce parasitic effects.

LIGHTNING AND AIRCRAFT

551.594.221:629.135



Thunderstorms are violent atmospheric phenomena, and on such a large scale that it hardly seems feasible to similate them in small-scale laboratory experiments.

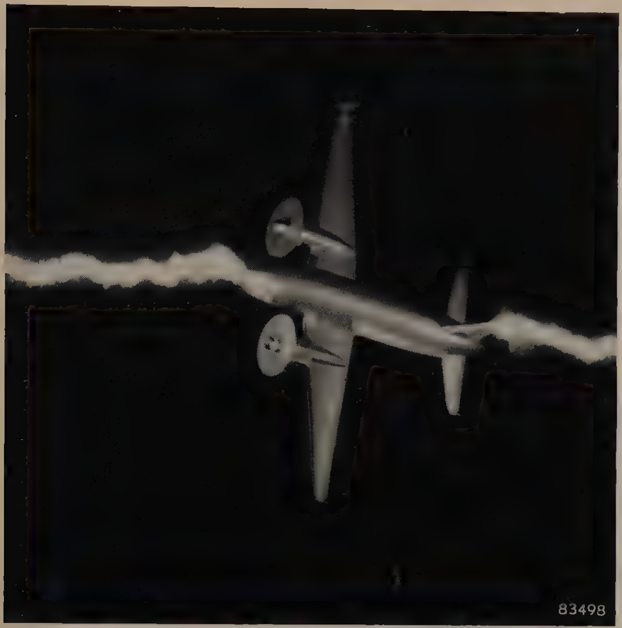
However, it is possible to study certain aspects of these phenomena by means of scale experiments, in those cases in which the effects are essentially unchanged by the scale. An example is the impulse testing of insulators and transformers at high tension, to simulate the effects of lightning striking a power line.

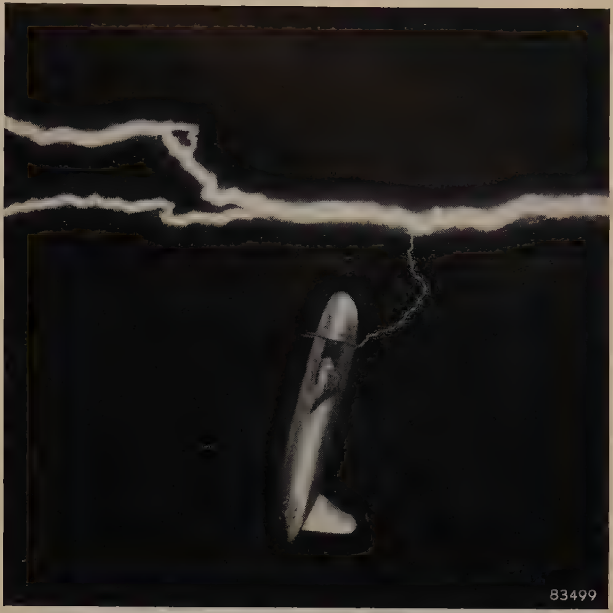
The photographs reproduced in this article are illustrative of quite a different investigation undertaken recently in Eindhoven. The problem is concerned with the question as to what parts of an aircraft flying through a thunderstorm region are most likely to be struck by lightning. With the increasing speed and size of modern aircraft, it has been found that incidents of this kind are becoming more and more frequent. Although accidents only rarely result from aircraft being struck by lighting, some damage, often only slight, usually results. In order to maintain full airworthiness the damage will probably have to be repaired before the next flight; this means loss of time which can have serious effects on the economics of operations.

To investigate which parts of an aircraft are most susceptible to a lightening stroke, a model aircraft, scale 1 : 100, was suspended from insulating threads between two high-tension electrodes (of dimensions of some yards, see photograph at the head of this article). One electrode is connected to earth, the other to the positive high-tension terminal of a million-volt cascade generator [1]). The separation of the electrodes is such that spark-over would occur spontaneously at any smaller separation. Only a faint brushing (corona discharge) occurs. This is caused by the inevitable roughness of the electrodes, dust particles on the surface, etc; brushing is also sometimes observed on the model aircraft. The total value of such corona currents is very small (a fraction of 1 mA). In the space between the

---

[1]) See e.g. A. Kuntke, Philips tech. Rev. **2**, 161, 1937; and A. Brouwers and F. A. Heyn, Philips tech. Rev. **6**, 46, 1941.





electrodes, around the model airplane, however, a critical condition of electric stress prevails, which may break down into a disruptive discharge by any slight change. A similar condition, on the verge of a discharge, may also prevail around an aircraft in a thunderstorm. A discharge can then be triggered by any adventitious factor causing a momentary increase in field strength at some point. As the experiments involved a large number of bolts of lightning — sufficient, that is, for a statistical evaluation of the results — it was obviously impracticable to wait each time for a chance disturbance. Instead, a disturbance was induced by shooting a pellet from an airgun against the high tension electrode. Once the voltage had been adjusted near enough to the critical value, virtually every shot would trigger a full discharge [2]).

Not all, but a large number of the discharges thus triggered, chose a path via the model aircraft. On one or two occasions the model was struck by a secondary discharge branching off from the main discharge (see photograph), a phenomenon which has also been observed with natural lightning. The place at which the model aircraft was struck by the discharge was photographically recorded. The aircraft model was placed in every conceivable spatial attitude with respect to the electric field. From the many hundreds of photographs is was possible to form some idea of the parts of the model most likely to be struck by the discharge.

On the basis of experiments of this kind it may well be possible in the future to provide certain points on aircraft with lightning arresters to prevent or minimize damage.

A. C. van DORSTEN.

---

[2]) The phenomenon that a rapid mechanical disturbance may cause a discharge across a gap when the applied voltage is just below the critical value was, as far as is known, first observed and studied by A. Kuntke in the Laboratory of C. H. F. Müller in Hamburg; see VDE-Fachberiche **12**, 157-164, 1948.

ENTROPY IN SCIENCE AND TECHNOLOGY

II. EXAMPLES AND APPLICATIONS

by J. D. FAST.

536.75

In a previous article [1]*) definitions and general aspects of entropy were treated; this article illustrates with examples the application of the entropy concept in chemistry, physics and engineering. As an example in the chemical field the reduction reactions of metal oxides, which play such an important part in metallurgy, are treated. In the field of physics, in which the statistical interpretation of entropy is explicitly expressed, the subject of paramagnetism is discussed, and the technique of paramagnetic demagnetization for the attainment of very low temperatures is explained. As examples of "classical" thermodynamics in engineering, hot-air engines, refrigerators and heat pumps are considered. The article is concluded with a discussion of the thermodynamical aspects of electromagnetic radiation.*

In the first article of this series (I) [1]) the concept of entropy has been examined from various viewpoints. In this article and in the two following ones we shall demonstrate with examples the importance of the entropy concept in widely divergent fields of chemistry, physics and technology.

These examples are taken more or less at random and are merely representative of the innumerable possible applications. Unfortunately restrictions of space do not permit us to enter into one of the finest examples, viz. accurately determining the equilibrium constant of reversible gaseous reactions (e.g. $H_2 + Br_2 \rightleftarrows 2\,HBr$) by the methods of statistical thermodynamics. We can only point out that the values of the entropy of polyatomic gases, on which such determinations are based, can be derived from their spectra. The various possible states of rotation and vibration of the gas molecules can be found from the wavelengths of the spectral lines. The values of the entropy can then be calculated from the available translational, rotational and vibrational states with the aid of the formula $S = k \ln m$ (cf. I)[2]).

Chemistry and metallurgy

In article I we have seen that a system whose temperature and volume are kept constant, strives towards a minimum value of the Helmholtz free energy

$$F = U - TS.$$

On the other hand, at constant temperature and *pressure*, the value of the Gibbs free energy (also known as free enthalpy and thermodynamic potential)

$$G = U + pV - TS = H - TS$$

strives towards a minimum.

Instead of the "competition" between the energy U and the entropy S, discussed under the heading *The free energy* in part I, here a similar thing occurs between the enthalpy H and the entropy S, viz. the striving of $H = U + pV$ towards a minimum and that of S towards a maximum value.

If a system of constant temperature and pressure contains different substances, then, according to the above, only those chemical reactions will occur spontaneously between the substances for which the change in Gibbs' free energy

$$\Delta G = \Delta H - T\Delta S \quad \ldots\ldots \quad \text{(II, 1)}$$

is negative.

In this equation ΔH is the heat of reaction at constant pressure; i.e. the amount of heat absorbed during an irreversible reaction at constant values of p and T, whilst ΔS stands for the increase in entropy as a result of the reaction.

The fact that ΔH is the ordinary (irreversible) heat of reaction at constant pressure, follows directly from the first law of thermodynamics, which can be stated here in the form:

$$Q_{\text{irr}} = \Delta U + p\Delta V = \Delta H. \quad . \quad . \quad \text{(II, 2)}$$

ΔU and ΔH, both representing the heat of reaction at constant pressure and constant volume, have nearly the same value as long as no gaseous components play a part in the reaction (e.g. $PbS + Fe \rightarrow Pb + FeS$), or if the number of gas molecules does not change during the reaction, e.g. ($H_2 + Cl_2 \rightarrow 2\,HCl$). If, on the other hand, the number of gas molecules formed during the reaction is not equal to the number of those that have disappeared (as in the reaction $2NH_3 \rightarrow N_2 + 3\,H_2$), then ΔU and ΔH will differ, according to (II, 2), by an amount

---

[1]) J. D. Fast, Entropy in science and technology I, Philips tech. Rev. **16**, 258-269, 1954/55, further to be referred to as I.

[2]) For an explanation of the methods used, cf. J. D. Fast, Entropie (in Dutch), Centen's, publishers, Amsterdam 1948, or R. W. Gurney, Introduction to statistical mechanics, McGraw Hill, New York, 1949.

$$p\Delta V = \Delta(pV) \approx n\,RT,$$

in which n stands for the difference between the number of gram- molecules of newly-formed gas and that of the disappeared gram-molecules of gas. In chemistry, ΔH is more important than ΔU, since most reactions take place at a constant pressure. According to the convention adopted in I, heat added to a system is designated positive (cf. I), so that an exothermic reaction has a negative heat of reaction.

When considered thermodynamically, it can be said that a chemical reaction will occur the more readily the more negative is the value of ΔG for this reaction. In thermodynamics, therefore, $-\Delta G = -\Delta H + T\Delta S$ is called the affinity of the reaction. This affinity increases if ΔH is more negative, i.e. if the reaction is more exothermic, and also if ΔS has a greater positive value, which means a greater increase in entropy in the course of the reaction.

Chemical reactions in which only solids participate usually involve only a very slight entropy change, and are consequently always exothermic if the reaction occurs spontaneously. If gases participate in the reaction, ΔS will be usually very small if the number of molecules remains unaltered during the reaction.

Conversely, if the number of gas molecules increases by the reaction, then generally the entropy will likewise increase substantially, because the number of available micro-states m in the gaseous state is far greater than in the condensed state (cf. I, in which it was shown that $S = k \ln m$).

Let us consider as an example the oxidation of carbon according to the two equations

$$C + O_2 \rightarrow CO_2, \quad . \ . \ . \ . \quad \text{(II, 3)}$$

$$2\,C + O_2 \rightarrow 2\,CO. \quad . \ . \ . \ . \quad \text{(II, 4)}$$

In the course of (II, 3) far more heat is evolved (the enthalpy decreases substantially more) than in the course of (II, 4). The entropy, on the other hand, hardly changes in the course of (II, 3), whereas during (II, 4), due to the doubling of the number of gas molecules, it increases considerably. The result of this is that at relatively low temperatures, where the energy (or enthalpy) term prevails, carbon oxidises to CO_2, whereas at high temperatures, at which the entropy term prevails due to the factor T (cf. equation II, 1), the combustion of carbon produces CO. In the latter case, according to (II, 1) (ΔS positive) ΔG will become increasingly negative at rising temperature, i.e. the affinity of carbon to oxygen, as regards the formation of CO, shows a continuous increase as the temperature rises.

Conversely, in the case of oxidation of solid or liquid metals into solid or liquid oxides, such as

$$2\,Fe + O_2 \rightarrow 2\,FeO, \quad . \ . \ . \ . \ . \quad \text{(II, 5)}$$

$$4\,Al + 3O_2 \rightarrow 2\,Al_2O_3, \quad . \ . \ . \ . \quad \text{(II, 6)}$$

the affinity decreases at rising temperature, since ΔS is negative due to the reduction of the number of gas molecules.

The fact that at rising temperature the affinity of some oxidation reactions (e.g. II, 5, or II, 6) diminishes, whilst that of others (reaction II, 4) increases, has the consequence that for every metal there is a temperature above which its affinity with oxygen at a certain pressure is less than that of carbon with oxygen. We have thus found an entropy effect that is especially important in metallurgy: at a sufficiently high temperature all liquid and solid metallic oxides can be reduced by carbon. For a more quantitative study we have to know the dependence upon the gas pressure. We shall not enter into the derivation, but only mention that for oxidation reactions of the type (II, 5) and (II, 6) at not excessively high O_2-pressures, we have the equation

$$\Delta G = \Delta G^0 - RT \ln p_{O_2} \, . \ . \ . \ . \ . \quad \text{(II, 7)}$$

where ΔG and ΔG^0 have a specific value for each reaction. ΔG^0 denotes the change of Gibbs free energy *at standard pressure*, and depends only on the temperature.

For reaction (II, 4) the equation (II, 7) becomes

$$\Delta G_{(4)} = \Delta G^0{}_{(4)} + RT \ln \frac{p^2_{CO}}{p_{O_2}} . \ . \ . \quad \text{(II, 8)}$$

In the equilibrium state $\Delta G = 0$, so for a simple oxidation reaction of the type (II, 5) and (II, 6), we can write from (II, 7),

$$\Delta G^0 = RT \ln p_{O_2} \, . \ . \ . \ . \ . \quad \text{(II, 9)}$$

Hence, from (II, 8), for reaction (II, 4) in equilibrium,

$$\Delta G^0{}_{(4)} = -RT \ln \frac{p^2_{CO}}{p_{O_2}} . \ . \ . \ . \quad \text{(II, 10)}$$

Knowledge of ΔG^0 at a certain temperature for e.g. reaction (II, 6) gives directly, according to (II, 9), the dissociation pressure of Al_2O_3 at that temperature, i.e. the oxygen pressure for which the affinity of reaction (II, 6) is zero. For oxygen at 1 atm, Al has an affinity of $-\Delta G = -\Delta G^0$, according to (II, 7). This is termed the "standard" affinity $-\Delta G^0$, which means the affinity when all substances that are formed or that disappear are at standard pressure. This implies in the present examples that all solids and liquids are in a pure state and that all gases have partial pressures of 1 atm. and are assumed to be perfect gases.

Fig. 1 shows the standard affinity of various oxidation reactions as a function of the temperature. All lines in the diagram relate to the reaction with 1 gram-molecule of oxygen. Starting in the lower left-hand corner we see that ΔG^0 for the reaction

$$4\,\mathrm{Ag} + \mathrm{O_2} \rightarrow 2\,\mathrm{Ag_2O} \quad \ldots\ldots \quad (\mathrm{II}, 1)$$

has a small negative value (and thus $-\Delta G^0$ has a small positive value). As explained above, the affinity decreases at rising temperature. This relationship is nearly linear, since ΔH^0 and ΔS^0 change only little with the temperature.

The slope of this "straight line" given by

$$\frac{\partial(-\Delta G^0)}{\partial T} = \Delta S^0 \quad \ldots\ldots \quad (\mathrm{II}, 12)$$

is a measure of the entropy change occurring in the reaction. The standard affinity of reaction (II, 11) reaches zero at approximately 200 °C, at which temperature Ag and $\mathrm{Ag_2O}$ are in equilibrium with oxygen at 1 atm: at that temperature the dissociation pressure of $\mathrm{Ag_2O}$ is 1 atm. Above 200 °C, at an $\mathrm{O_2}$-pressure of 1 atm., $\mathrm{Ag_2O}$ dissociates into silver and oxygen.

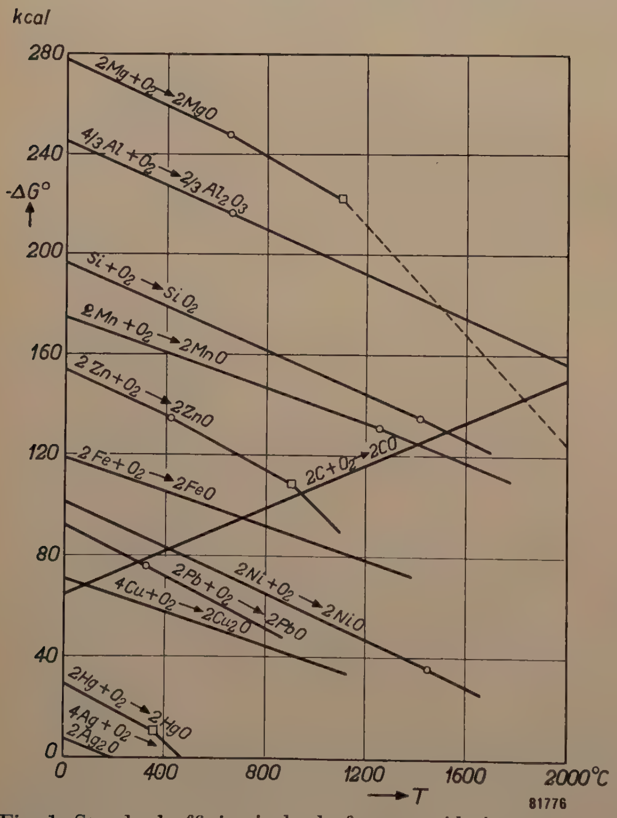


Fig. 1. Standard affinity in kcal of some oxidation reactions plotted as functions of the temperature. The circles in the diagram corresponds to the melting points, and the squares to the boiling points of the various metals. The affinity lines have not been produced beyond the melting points and boiling points of the oxides.

At the point where the metal or the oxide undergoes a change of state, there is a discontinuity in the affinity curve. A number of the lines in the diagram show this feature, e.g. the second from the bottom, relating to the reaction

$$2\,\mathrm{Hg} + \mathrm{O_2} \rightarrow 2\,\mathrm{HgO}. \quad \ldots\ldots \quad (\mathrm{II}, 13)$$

The discontinuity in this case occurs at 357 °C, the boiling point of mercury. Above this temperature, mercury at 1 atm. is in the gaseous state; its entropy is then far greater than in the liquid state. $\Delta S^0{}_{(13)}$ is therefore more negative above 357 °C than below this temperature and the curve shows a steeper slope. Melting points and crystallographic transition points, on the other hand, cause only slight changes in the direction of the affinity curves, due to their relatively small entropy changes. For the solid and the liquid state of metal and oxide all the lines have roughly the same slope. This is due to the fact that the entropy change of all these reactions is mainly determined by the disappearance of 1 molecule of $\mathrm{O_2}$.

From a metallurgical point of view, the most important exception to these roughly equal slopes is the curve for the reaction $2\mathrm{C} + \mathrm{O_2} \rightarrow 2\,\mathrm{CO}$, which rises with temperature as already mentioned. Wherever it intersects another line, e.g. that of

$$\mathrm{Si} + \mathrm{O_2} \rightarrow \mathrm{SiO_2} \quad \ldots\ldots \quad (\mathrm{II}, 14)$$

the standard affinities of the reactions (II, 4) and (II, 14) are equal, which means that the standard affinity of the reaction

$$\mathrm{Si} + 2\,\mathrm{CO} \rightarrow \mathrm{SiO_2} + 2\,\mathrm{C}, \quad \ldots \quad (\mathrm{II}, 15)$$

obtained by subtracting (II, 4) from (II, 14), is zero at the temperature of the point of intersection. This means that the CO equilibrium pressure of this reaction, and hence also of the inverse reaction

$$\mathrm{SiO_2} + 2\,\mathrm{C} \rightarrow \mathrm{Si} + 2\,\mathrm{CO}\,. \quad \ldots \quad (\mathrm{II}, 16)$$

is 1 atm. at this temperature. In fact, this pressure is reached at a lower temperature, as there exists rather a large affinity between Si and C, because of which the reaction with carbon in reality progresses further, according to:

$$\mathrm{SiO_2} + 3\,\mathrm{C} \rightarrow \mathrm{SiC} + 2\,\mathrm{CO}. \quad \ldots \quad (\mathrm{II}, 17)$$

This reaction corresponds to higher CO-pressures than reaction (II, 16). Also the formation of gaseous SiO should be reckoned with, which likewise causes the interaction between C and SiO_2 to occur at lower temperatures than would be expected from (II, 16).

It should be mentioned that the affinity values do not give any information on the *speed* of a chemical reaction. There are cases in which ΔG is strongly negative, yet the speed of reaction is extremely slow. Fig. 1, for example, indicates that the standard affinity of reaction (II, 6) is very great at normal room temperature, whereas it is known from experience that aluminium objects are not appreciably corroded by oxygen. Further examination shows that a very thin coating of oxide has rapidly formed on the metal, thus protecting it from further attack, since neither Al, nor 0-atoms (or ions) are capable of penetrating this skin.

The investigation of the speeds of chemical reactions falls outside the scope of thermodynamics. The latter is concerned only with states of equilibrium. The great value of thermodynamics in chemistry however, is that it can be established whether a given reaction will proceed in the desired direction simply by checking that its ΔG is negative. Of equal importance are the quantitative conclusions which may be drawn, viz. the chemical equilibrium of a reaction can be derived from ΔG^0; i.e. the maximum possible yield of reaction products can be calculated in advance. As an example, let us consider reaction (II, 4), for which we can put, using equations (II, 1) and (II, 10):

$$\ln \frac{p^2_{CO}}{p_{O_2}} = -\frac{\Delta H^0}{4.575\,T} + \frac{\Delta S^0}{4.575}. \qquad \text{(II, 18)}$$

Instead of calculating the equilibrium directly from ΔG^0, which is difficult to be determined directly, it can also be found from ΔH^0 and ΔS^0. The former of the two, the constant-pressure heat of reaction under standard pressure, is known from calorimetric measurements. The latter (ΔS^0), the entropy change at standard pressure can be calculated from the specific-heat values which are known from calorimetric and spectroscopic data.

The particular importance of the knowledge of thermodynamics to the metallurgist was strikingly demonstrated in the past century by the great French chemist Le Chartelier. His line of reasoning was as follows. It is known that in blast-furnaces the reduction of iron oxides is mainly effected by their reaction with CO, with the formation of CO_2. The gas leaving the blast-furnace at the top, however, contains a considerable percentage of CO, which was considered wasteful and undesirable. Since the incompleteness of the reaction was attributed to the insufficiently prolonged contact between CO and iron ore, blast-furnaces of abnormal heights were errected. The percentage of CO in the escaping gas, however, did not diminish. These costly experiments showed that there is a certain upper limit to the efficiency of CO as a reducing agent for iron oxides, which cannot be exceeded. Had the laws governing the chemical equilibrium been known, the same conclusion could have been reached far more quickly and at far less cost. It is only necessary to know the reaction enthalpy ΔH^0 and the reaction-entropy ΔS^0 of the reaction

$$FeO + CO \rightarrow Fe + CO_2 \quad . \; . \; . \quad \text{(II, 19)}$$

for the standard conditions, to be able to calculate directly the maximum fraction of the CO that can be converted into CO_2 from the equation

$$\ln \frac{p_{CO_2}}{p_{CO}} = -\frac{\Delta H^0}{4.575\,T} + \frac{\Delta S^0}{4.575} \quad . \; \text{(II, 20)}$$

Reaction (II, 19) invloves only relatively small changes in enthalpy and entropy, viz. at 25 °C $\Delta H^0 = -3800$ cal and $\Delta S^0 = -3.0$ cal/degree. With the aid of these values, and disregarding their temperature dependence, it is found by (II, 20) that escaping ($CO + CO_2$)-mixture will have at least 40%-60% of CO in the temperature range 1000-2000 °K [3]).

Paramagnetism

The molecules of paramagnetic substances possess a magnetic moment, i.e. they behave to a certain extent as small magnets. The property of paramagnetism is a consequence of the movements of the electrons in the molecule as well as of the angular momentum (spin) of the electrons. In most molecules the motions of the electrons are coupled in such a way that the resulting magnetic moment is zero.

An example of a gaseous paramagnetic substance is oxygen. The magnetic properties of O_2 are based on the fact that two of the electrons of each oxygen molecule are "unpaired", i.e. have similarly directed (i.e. parallel) angular momenta (spins) and consequently also parallel magnetic moments of the same sign. All the other electrons in the O_2 molecule, on the other hand, are paired. A gaseous alkali metal provides an even simpler example, because here the paramagnetic properties are based on the presence of only one non-compensated electron spin per atom of the gas. Several hydrated salts show a marked analogy to these paramagnetic gases, because the spin carriers in these salts are separated by such a distance that their interaction, just as in the gases, may be neglected to a first approximation.

---

[3]) A more accurate calculation, taking into account the temperature-dependence of ΔH° and ΔS°, gives even higher percentages of CO.

If oxygen, gaseous sodium or a paramagnetic salt, is brought into a magnetic field, then the striving towards a minimum value of the energy corresponds to the striving of the permanent atomic magnets to assume an orientation with their axes parallel to the field, just as compass needle in the magnetic field of the earth. A parallel orientation of all magnetic axes is by no means attained, at least not at high temperature, since the thermal agitation associated with the striving of the entropy towards a maximum value is opposed to this. The orientation of each particle with respect to the direction of the field is continuously changing due to the thermal agitation but, averaged over all molecules, there always remains a small resultant moment in the direction of the field.

In view the two opposed tendencies — the energy and the entropy effect — it is clear that the resultant moment (the magnetization) will be related directly to the field strength H but inversely to the absolute temperature T. In fact, it is found to be proportional to H and inversely proportional to T. The dependence on the temperature is known as Curie's law.

If the field strength is further increased, the magnetization is, in the long run, unable to keep up the same rate of increase. Finally a state of saturation is bound to occur in which all magnetic axes are in the parallel orientation. Fields strong enough to attain this saturation at normal room temperature, however, cannot be generated in the laboratory but by applying a combination of very strong fields and very low temperatures this state of saturation has been very slosely approached for several paramagnetic salts. The same principle has also been employed to reach temperatures considerably lower than hitherto possible. Before we describe this in the following section, we shall derive first the dependence of the magnetization on H and T for a simple case.

We shall consider the case that the paramagnetic properties are based on the presence of one non-compensated electron spin per molecule [4]) and that the directions of the magnetic moments are practically independent of each other. The quantum theory states that the component of the spin in the direction of the magnetic field can assume only the values $+\frac{1}{2}$ and $-\frac{1}{2}$ (expressed in units of $h/2\pi$). This corresponds to a magnetic moment of $+\mu_B$ or $-\mu_B$ in the direction of the field, if the symbol μ_B represents a Bohr magneton. The energy levels of these two states coincide when the field strength approaches zero. At a field strength H, however, the energy levels relating to the field-free state have the values $-\mu_B H$ and $+\mu_B H$. In this case, then, we are concerned with only two energy levels. The distribution of the magnets between these levels as a function of H and T can be derived as follows.

In a system of N molecules, at a given value of H a larger fraction will be formed at the lower than at the higher energy level. If the difference between the two populations is denoted by n, then $\frac{1}{2}(N+n)$ molecules have their magnetic moments in the direction of the field and $\frac{1}{2}(N-n)$ in the opposite direction. The system then possesses a moment $n\mu_B$ in the direction of the field and an energy

$$U = U_0 - \tfrac{1}{2}(N+n)\,\mu_B H + \tfrac{1}{2}(N-n)\,\mu_B H = U_0 - n\mu_B H, \qquad \text{(II, 21)}$$

where U_0 represents the energy in the absence of the magnetic field. According to I, the entropy of the system is determined by the number of possible distributions of N molecules, subject to the condition that a fraction $\frac{1}{2}(N+n)$ is in one state and a fraction of $\frac{1}{2}(N-n)$ in the other state:

$$S = S_0 + k \ln \frac{N!}{\left\{\frac{1}{2}(N+n)\right\}!\,\left\{\frac{1}{2}(N-n)\right\}!}. \qquad \text{(II, 22)}$$

In this equation S_0 represents that part of the entropy which is independent of the orientation of the moments.

The free energy $F = U - TS$ contains, therefore apart from a term $U_0 - TS_0$ independent of n, also a term that can be approximated by means of Stirling's formula (I, 3) by

$$F(n) = -n\mu_B H + kT\left\{\frac{N+n}{2}\ln\frac{N+n}{2N} + \frac{N-n}{2}\ln\frac{N-n}{2N}\right\}. \qquad \text{(II, 23)}$$

If the system could submit completely to the tendency towards a minimum value of its energy, all molecules would be at the lowest level ($n = N$). If, on the other hand, it could submit to the tendency towards a maximum entropy value, then the molecules would be equally divided between the two levels ($n = 0$). The compromise (the equilibrium state) lies according to I at the point where the free energy is minimum, i.e. where $\mathrm{d}F(n)/\mathrm{d}n = 0$.

---

[4]) For the sake of convenience the term "molecule" is used here. Clearly the discussion also applies to atoms and ions.

Hence $$\frac{kT}{2}\ln\frac{N+n}{N-n} = \mu_B H$$

or, after re-arranging,

$$\frac{n}{N} = \frac{e^{\mu_B H/kT} - e^{-\mu_B H/kT}}{e^{\mu_B H/kT} + e^{-\mu_B H/kT}} = \tanh(\mu_B H/kT).$$

The magnetic moment of the system is, therefore,

$$n\mu_B = N\mu_B \tanh(\mu_B H/kT).$$

The moment per unit volume is

$$I = N'\mu_B \tanh(\mu_B H/kT). \quad . \; . \; . \; (\text{II}, 24)$$

where N' represents the number of molecules per unit volume. In case of $\mu_B H/kT \ll 1$, i.e. where relatively weak fields or relatively high temperatures are concerned, $\tanh(\mu_B H/kT) \approx \mu_B H/kT$, and consequently

$$I \approx \frac{N'\mu_B^2}{kT} H. \quad . \; . \; . \; . \; (\text{II}, 25)$$

For this limiting case we have thus arrived at the relationship between I and H and between I and $1/T$, mentioned earlier.

The susceptibility χ in this region is given by

$$\chi = \frac{I}{H} = \frac{N'\mu_B^2}{kT}. \quad . \; . \; . \; . \; (\text{II}, 26)$$

If a molecule possesses more than one non-compensated electron-spin, these spins often combine into one resultant molecular spin. Such a resultant spin has more than two possible orientations in a magnetic field.

For a spin quantum number s ($s = \frac{1}{2}$ for one non-compensated electron spin) the general rule is that the component in the direction of the magnetic field can assume only the values $s, s-1, s-2, \ldots, -s+1, -s$. For the case under discussion ($s = \frac{1}{2}$) the components $+\frac{1}{2}$ and $-\frac{1}{2}$ were possible. For $s = 1$ (two similar spin orientations per molecule) we find the three possibilities $+1$, 0 and -1. As a rule there are $2s + 1$ different possibilities of orientation of the resultant spin.

A calculation analogous to the one carried out for $s = \frac{1}{2}$, shows that for $\mu_B H/kT \ll 1$,

$$\chi = \frac{N' 4s(s+1)\mu_B^2}{3\,kT}, \quad . \; . \; . \; (\text{II}, 27)$$

which reduces to (I, 26) for $s = \frac{1}{2}$.

Formula (II, 27) is frequently written as

$$\chi = \frac{N'\mu_{\text{eff}}^2}{3\,kT}, \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; (\text{II}, 28)$$

where μ_{eff} represents the effective magnetic moment defined by:

$$\mu_{\text{eff}}^2 = 4s(s+1)\,\mu_B{}^2.$$

Equation (II, 28) is identical in form to the equation for the contribution of a dipole to the electric susceptibility, χ_{dip}, of a rarefied gas of molecules such as H_2O, NH_3, HCl, which possess a permanent electric moment μ_{el}:

$$\chi_{\text{dip}} = \frac{N'\mu_{\text{el}}^2}{3\,kT}. \quad . \; . \; . \; . \; . \; . \; . \; (\text{II}, 29)$$

The theoretical background of the formulae (II, 28) and (II, 29), however, is somewhat different. In very strong fields or at extremely low temperatures the electric polarization is given by $N'\mu_{\text{el}}$, whereas the magnetization is *not* defined by the analogous expression $N'\mu_{\text{eff}} = N'2\sqrt{s(s+1)}\,\mu_B$, but by $N'2s\mu_B$.

Fig. 2 gives the curves of the magnetic moment as a function of H/T for three paramagnetic salts. Consider the middle curve, which refers to $FeNH_4(SO_4)_2{\cdot}12H_2O$, i.e. ferric ammonium alum. In this compound the ferric ion embodies the magnetic properties. In its outer electron shell it contains $2 + 6 + 5$ electrons ($3s^2\,3p^6\,3d^5$). Of these, the five $3d$-electrons are unpaired, i.e. they have parallel spins. All other electrons are paired and cannot be influenced by a magnetic field. The spin quantum number of this ion is hence $s = 5/2$. In accordance with this, the curve shows that the mean magnetic moment in the direction of the field, at large values of H/T, reaches a saturation value of 5 Bohr

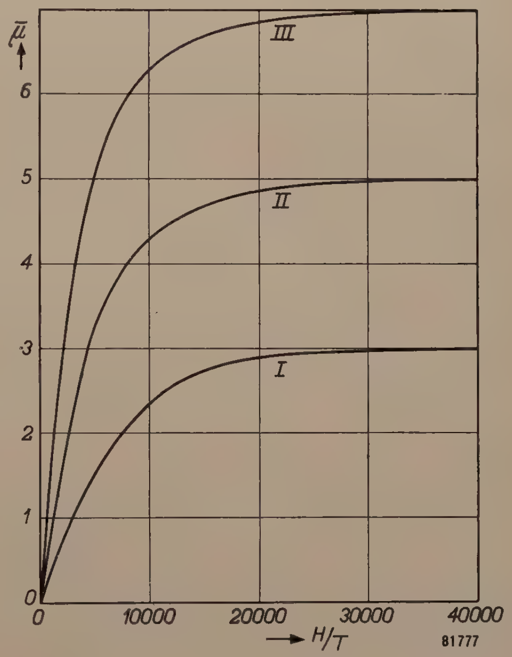


Fig. 2. Average magnetic moment $\bar{\mu}$ in Bohr magnetons per molecule as functions of H/T in oersteds per °K for (*I*) sodium-chrome alum ($s = 3/2$), (*II*) ferric-ammonium alum ($s = 5/2$) and (*III*) gadolinium sulphate octahydrate ($s = 7/2$). From measurements of C. J. Gorter, W. J. de Haas and J. van den Handel, Proc. Kon. Ned. Akad. Wetensch. **36**, 158-167, 1933, and W. G. Henry, Phys. Rev. **88**, 559-562, 1952.

magnetons per ion. The curve shows also that this mean moment in the direction of the field is proportional to H/T at small values of the latter. The two other curves refer to chrome alum (s = 3/2) and gadolinium sulphate-octahydrate (s = 7/2).

Attaining very low temperatures

By allowing liquid helium to boil under reduced pressure, a temperature not much lower than about 1 °K can be attained. Relatively much lower temperatures may be reached by employing the paramagnetic properties of a salt such as the above-mentioned ferric-ammonium alum. Pioneer work in this field was carried out in the Netherlands in the Kamerling-Onnes Laboratory of Leiden University [5]).

We have shown in the above that the resultant of five parallel spins and hence also the resultant spin of a ferric ion can assume only six different orientations relative to a magnetic field. If the field is very weak, there is little preference energetically for any of the six orientations, i.e. the six energy levels corresponding to the possible orientations are almost coincident in a weak field, and at not too low temperatures they will be equally occupied.

At small values of H/T, therefore, each molecule of the ferric-ammonium alum may occur in any one of six different but equally probable paramagnetic states. For each of the six possible states of the first molecule, the second molecule may be present in any one of the six states. There are thus 6^2 different states for two molecules, and 6^{N_0} different states for a gram-molecule (N_0 molecules). That part of the entropy dependent on the paramagnetic properties of the salt, (also called paramagnetic entropy, or, more specifically, spin entropy), has, according to the well-known formula $S = k \ln m$, the value

$$S = k \ln 6^{N_0} = R \ln 6$$

per gram-molecule, R being the gas constant. To put it more generally: if s is the quantum number of the resultant spin, then the spin entropy per gram-molecule is given by

$$S = R \ln (2s + 1) \quad \ldots\ldots\ldots \quad \text{(II, 30)}$$

The energy levels of the (2s + 1) different states become more separated as the magnetic field becomes stronger. At a very strong field and a continuously dropping temperature the higher energy levels will be gradually emptied until ultimately the state of saturation, discussed earlier, is approached, in which practically all molecules are at the lowest level. The paramagnetic entropy per gram-molecule for this extreme case is given by

$$S \approx R \ln 1 = 0.$$

A considerable lowering of the temperature can now be attained by bringing about an adiabatic and reversible demagnetization of the magnetically saturated substance. The effect of this is completely analogous to the temperature drop occurring with the adiabatic -reversible expansion of a gas. In either case the temperature drop can be achieved by a two-step procedure. First stage: the volume of the gas is reduced in an isothermal (reversible) way, or the magnetization of the salt is isothermally and reversibly increased. In either case the result is a decrease of the entropy (in the case of the gas, as in I, p. 209), which means (since $T\mathrm{d}S = \mathrm{d}Q_{\mathrm{rev}}$) a flow of heat out of the gas or the salt. This is carried off to a heat sink at constant temperature (a quantity of liquid helium, boiling under reduced pressure in the case of the paramagnetism). Second stage: the thermal contact with the surroundings is broken, after which the pressure on the gas is reversibly reduced, or the magnetic field is reversibly decreased to zero. During this adiabatic-reversible process the total entropy of the gas or salt, since $\mathrm{d}S = \mathrm{d}Q_{\mathrm{rev}}/T$ and $\mathrm{d}Q_{\mathrm{rev}} = 0$, remains unaltered.

In the case of the gas, it can be seen directly that the increase of the entropy due to the expansion has to be compensated by a reduction of the thermal entropy. This implies a decrease of the temperature [6])
The case of the paramagnetic salt is not quite so simple. An analogous reasoning would be as follows. When the magnetic field is reduced to zero, the paramagnetic entropy $S = R \ln (2s + 1)$ increases and this happens at the cost of the vibrational entropy of the atoms, since the total entropy remains constant. In fact, this latter entropy is extremely small at an initial temperature of approximately 1 °K, far smaller than the paramagnetic entropy. At first sight it could thus be expected that during the adiabatic demagnetization the temperature would drop to 0 °K, whilst, moreover, a considerable part of the magnetic order would be retained. The latter is indeed the case, but the temperature does not drop to 0 °K. The lowest temperatures that can be

---

[5]) See, e.g., W. J. de Haas, E. C. Wiersma and H. A. Kramers, Physica **1**, 1-13, 1934.

[6]) That this adiabatic-reversible process is extremely suitable for producing low temperatures is demonstrated by the development of the gas refrigerating machine, which operates on this principle. See the articles by G. W. L. Köhler and C. O. Jonkers, Philips tech. Rev. **16**, 69-78, 1954/55 (No. 3) and **16**, 105-115, 1954/55 (No. 4).

attained in this way are between 0.01 and 0.001 °K. That the temperature cannot be reduced further is due to a slight interaction among the magnetic moments and between the moments and the crystal lattice, which have not been taken into account up to now. These interactions have the effect that the energy levels of the (2s + 1) possible states of the paramagnetic ions do not coincide completely, even in the absence of an external magnetic field. In accordance with the third law of thermodynamics (cf. I) this causes all spins to have a fixed orientation at absolute zero. This means that at very low temperatures and in the absence of a magnetic field, the disorder of the spins automatically disappears under the influence of the interaction, causing the entropy to drop to zero. The curve of *fig. 3* represents the entropy of a gram-molecule of ferric-ammonium alum (in the absence of an external magnetic field) as a function of the temperature. Between 0.5 °K and 5 °K the entropy remains nearly constant at the earlier mentioned value of $R \ln 6$. The effect of the lattice vibrations on the entropy only becomes discernible above 5 °K.

In the preceding section we have seen that the difference between two energy levels of an electron in a field H is $2\,\mu_B H$. To a first approximation, we may assume that the (2s + 1) levels of the paramagnetic ions are separated by uniform distances $2\mu_B H$ in a strong field H, and that they are situated at far smaller intervals $k\Theta_0$ in a zero field (see *fig. 4*). Furthermore, (see above) we may disregard the entropy of the lattice vibrations at 1 °K or lower with respect to the paramagnetic entropy. The latter therefore, remains nearly constant during the adiabatic demagnetization, which means in the statistical concept of entropy that the electron spins at the final temperature T_2 (after removal of the field) are distributed among the right-hand group of levels in fig. 4 in the same arrangement as they were at the initial temperature T_1 in the field H among the middle group. According to I this distribution is determined by the ratio of the separation between the levels to kT, i.e. by $2\mu_B H/kT$ and $k\Theta_0/kT_2$ respectively. To a rough approximation, therefore,

$$\frac{2\mu_B H}{kT_1} \approx \frac{k\Theta_0}{kT_2}, \quad \text{or} \quad T_2 \approx \frac{k\Theta_0}{2\,\mu_B H}\,T_1. \qquad \text{(II, 31)}$$

The final temperature T_2 is thus lower as the interaction between the magnetic moments (and hence $k\Theta_0$) is less, the field H stronger, and the starting temperature T_1 lower. The first condition is satisfied by using salts such as $FeNH_4(SO_4)_2 \cdot 12H_2O$ in which the paramagnetic ions (Fe^{+++}) lie far apart. It is to be expected that in future even lower temperatures can be attained by employing the very weak coupling of the spins of certain atomic nuclei.

The foregoing shows that below 1 °K the concept of temperature is not so much coupled to the vibrational disorder but to the disorder of the spins of the system. The emptying of the higher energy levels is already perceptible at a temperature for

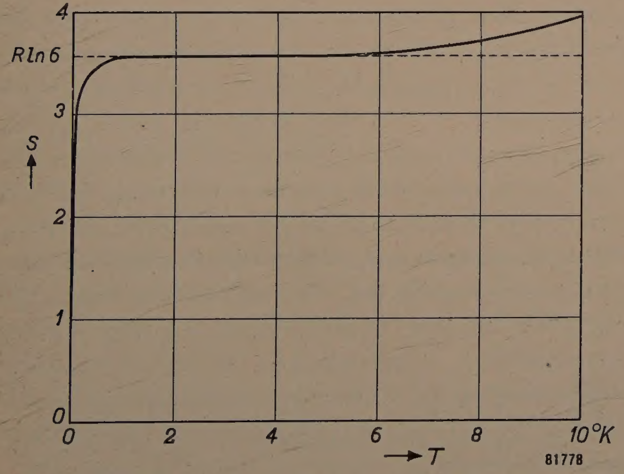


Fig. 3. Entropy in cal per °K per gram-molecule of ferric-ammonium alum (in the absence of a magnetic field) as a function of the temperature.

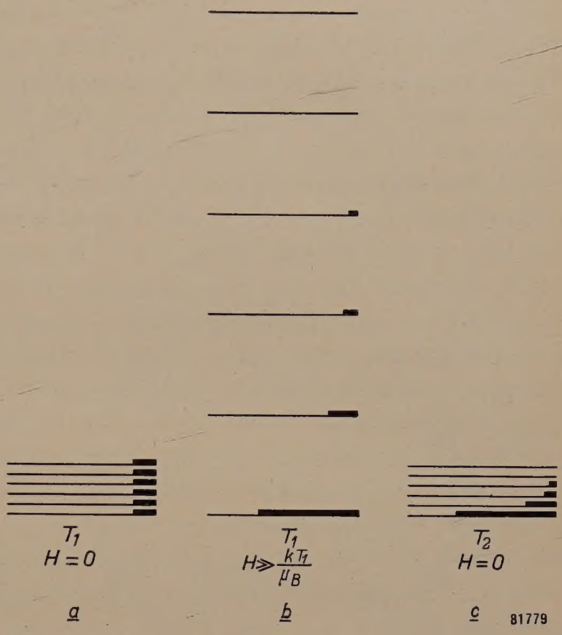


Fig. 4. Adiabatic demagnetization.
a) At the initial temperature T, kT_1 is considerably greater than the separation of the energy levels; the spins are uniformly distributed among the six levels. This is indicated by the thick parts of the lines.
b) When a strong field such that $\mu_B H \gg kT$, is applied isothermally, the higher levels will be emptied.
c) When the field is adiabatically removed, the entropy remains constant, which means that although the levels come as near together as in the beginning, the distributions remain the same as in (*b*). (From an article by N. Kurti in the booklet "Low temperature physics", Pergamon Press, London 1952.

which kT is equal to the separation of the levels, i.e. at the temperature $T = \Theta_0$. This "characteristic" temperature may be considered as a Curie-temperature. Its approximate value can be derived from specific-heat measurements.

Heat engines

Considerations of entropy play an important part in the study of machines which convert heat into work, which may be collectively termed heat engines (steam engines, hot-gas engines, etc.). The motions of the atoms of a body, e.g. the piston of a steam engine, only provide useful work when they are uniformly directed, i.e. when the piston moves as a whole. The conversion of heat into work means in principle the conversion of the completely random motions of atoms into a uniformly directed (ordered) motion. i.e. a decrease of the entropy. According to the second law of thermodynamics this is not feasible unless compensated in some way. Indeed, one of the earliest formulations of the second law of thermodynamics was: It is impossible to abstract heat from a heat-reservoir and to convert it *completely* into work.

The reverse process, on the other hand, the complete conversion of work into heat (e.g. by friction which occurs spontaneously and therefore irreversibly, is a commonplace. This is sometimes referred to as the "degeneration of energy", since here a form of energy valuable to mankind is changed into a less useful form.

How do we succeed, in spite of all this, in converting heat into work? By using the striving of the free energy towards a minimum value, discussed in the first article, it is possible to decrease the entropy, provided that this decrease is compensated by a simultaneous decrease of the energy of the (non-isolated) system. This method cannot be used in practice, however, because a heat engine has to complete the same cycle again and again, whereby at the end of each cycle the piston returns to its initial position and the initernal energy reaches the initial value again.

The requirement of the second law has therefore to be met in another way, viz. by the fact that only part of the heat of combustion of the fuel is converted into work, while another part is carried off to a cooler, i.e. to a heat sink at a lower temperature level. The reduction of the entropy, corresponding to the conversion of heat into mechanical energy of the piston, is thus offset by an entropy increase caused by the flow of another quantity of heat from the engine to the cooler.

Taking the steam engine as an example, its working cycle is as follows. The steam absorbs a quantity of heat Q_1 at a high temperature T_1 from the boiler, and then discharges a smaller absolute heat value Q_2 at a lower temperature T_2 to the condenser. The difference $Q_1 - Q_2$ is gained in the form of work.

According to the second law in its form $\mathrm{d}Q/T \leqq \mathrm{d}S$, which becomes $\oint \mathrm{d}Q/T \leqq 0$ for a cyclic process, the following equations applies to our simplified model of a steam engine:

$$\frac{Q_1}{T_1} - \frac{Q_2}{T_2} \leqq 0,$$

or:

$$\frac{Q_1 - Q_2}{Q_1} \leqq \frac{T_1 - T_2}{T_1}. \quad . \; . \; . \; (\mathrm{II}, 32)$$

The first term of this equation simply represents the proportion of the applied heat that has been converted into work, i.e. the efficiency of the heat engine. The formula indicates that the highest possible efficiency would be achieved by a machine with completely reversible cycle (Carnot cycle); this optimum efficiency would then equal $(T_1 - T_2)/T_2$.

For many types of steam engines we may take $T_1 \approx 430$ °K and $T_2 \approx 300$ °K. The optimum efficiency is therefore about 0.3 (30 %). In practice the efficiency of this type of machine remains below 20 %.

Refrigerators and heat-pumps

The heat engine discussed above can also be made to function in the reverse direction. Instead of displacing heat from a reservoir of high temperature to one of low temperature and converting the greatest possible portion of it into work "on the way", we now perform work in order to transfer heat from the reservoir at low temperature to that at high temperature. Whereas with the former cycle the working fluid (the gas or the vapour) took up more heat at high temperature than it discharged at low temperature, it will now discharge more heat at high temperature than it absorbs at low temperature The difference constitutes the minimum external energy W that is required for driving the machine. Such a machine is termed a refrigerator if its function consists primarily of abstracting heat from the reservoir at low temperature; if its task consists essentially of delivering heat to the reservoir at high temperature it is termed a heat-pump. In the former case we are interested in Q_2, and the efficiency is given by:

$$\frac{Q_2}{W} \leqq \frac{T_2}{T_1 - T_2}. \quad . \; . \; . \; . \; . \; . \; (\mathrm{II}, 33)$$

For a heat-pump the essential factor is Q_1 and the efficiency is given by

$$\frac{Q_1}{W} \leq \frac{T_1}{T_1-T_2} \quad . \; . \; . \; . \; . \; . \; (\mathrm{II}, 34)$$

Let us illustrate our point by considering the case that $T_1 = 300$ °K and $T_2 = 270$ °K. When our machine is used as a domestic refrigerator, it has the purpose of keeping the interior of the box at 270 °K = —3 °C, whilst the ambient temperature is 300 °K = 27 °C. When functioning as a heat-pump it has the task of abstracting heat from outside at —3 °C and using it for keeping the interior of a building at 27 °C. The upper limit of the "efficiency" (or better, the *coefficient of performance*) is 270/30 = 9 for the refrigerator and 300/30 = 10 for the heat-pump. In practice, of course, the efficiency remains considerably below these values 7)).

Compared with a heat pump, heating by means of an electric resistance heater is very uneconomical. At first sight one is inclined to think that there is hardly a more favourable proposition than direct electric heating, since all the energy is converted into heat. The foregoing, however, shows that a heat pump may be several hundred percent more "efficient". The heat pump is as yet used only to a very limited extent. In practice it commonly takes the form of a liquid with high latent heat of evaporation (e.g. CF_2Cl_2) which is made to evaporate at low temperature, during which it abstracts the necessary heat from a source at low temperature, e.g. the outside atmosphere or the ground. Condensation is then made to take place at the higher temperature, delivering the heat of condensation to where it is required e.g. the interior of a building. It can be safely predicted that as conventional sources of energy become more and more exhausted, interest in the heat pump will increase. This machine is particularly suited to the heating of buildings, because its efficiency is especially high for small values of $T_1 - T_2$ (see eq. II, 34).

Radiation of heat and light

All bodies radiate energy, or, in other words, emit photons. The quantity of energy emitted per unit time and per unit area increases rapidly with the temperature and depends on the nature of the radiating body. The "blacker" the body, i.e. the more radiation it absorbs, the better emitter of radiation it will be. Maximum emission occurs with the "perfect black body", the body that absorbs all incident light. Such a body does not exist, but can be very closely approximated by a small aperture in the wall of an otherwise light-tight chamber. Any radiation entering through the hole is scattered and absorbed by repeated reflection (*fig.* 5), so that only an infinitesimal fraction escapes. If the chamber is at a temperature T, the aperture radiates as a perfect black body of that temperature. What is emitted is the same radiation that is present inside the chamber.

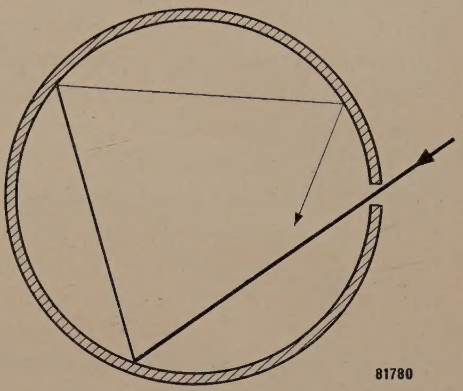


Fig. 5. A small aperture in a hollow body acts as a perfect black body, i.e. a perfect emitter and absorber of radiation.

We shall not concern ourselves here with the distribution of the radiation among the different wavelengths, but merely use thermodynamical arguments to arrive at the Stefan-Boltzmann law for the total radiation of a black body.

Consider a hollow chamber in the form of a cylinder provided with a weightless piston that moves without friction. The bottom or another part of the vessel is a black body, all other walls, and the face of the piston, being perfectly reflecting. The black end of the cylinder continuously emits and absorbs radiation. Since the radiation has a finite speed of propagation, there is always a quantity of radiation energy U present inside the hollow chamber. This chamber filled with radiation is in several respects analogous to a gas-filled vessel. An important difference, however, is that the density of the radiation energy $u = U/V$ depends exclusively on the temperature T, whereas the density of the gas can only be altered by altering the volume. The total number of photons in a hollow chamber changes with the temperature; the total number of gas molecules remains constant.

The pressure exerted by the radiation (photons) on the wall of our (gas-free) chamber is given by $p = \frac{1}{3}\mu$, according to electromagnetic theory.

---

7) It is usually preferable to indicate the performance of a refrigerator or heat pump by a quantity that cannot exceed unity. The concept then introduced is the *figure of merit*, this being the ratio of the actual to the theoretical (Carnot) efficiency. In this connection, see the passage in small print on p. 73 of the first article quoted under 6), and fig. 2 of the second article.

Work can be done on the system by compressing the radiation "gas" by displacing the piston, and heat can be supplied to it by bringing the black end of the cylinder in contact with a source at higher temperature. Because $p = \frac{1}{3}u$, the first law takes the form

$$\mathrm{d}U = \mathrm{d}Q - \tfrac{1}{3}u\,\mathrm{d}V.$$

The radiation in the chamber also possesses a certain entropy. This statement is statistically justified by considering the radiation as a gas composed of photons, each of which can occur in different energy states. It may be justified in an even simpler way by classical thermodynamics: if the heat source delivers a quantity of heat dQ reversibly to the cylinder, the entropy of the source drops. Since the total entropy must remain constant with a reversible transfer, the entropy S of the radiation (the photon-gas), must increase as a result of the transfer. The combination of the first and second law of thermodynamics thus gives:

$$\mathrm{d}U = T\mathrm{d}S - \tfrac{1}{3}u\,\mathrm{d}V. \quad . \quad . \quad . \quad (\text{II}, 35)$$

where T is the temperature of the black-body radiation, i.e. the temperature of the black wall with which the radiation is in equilibrium. Employing the relation $U = Vu$ or $\mathrm{d}U = V\mathrm{d}u + u\mathrm{d}V$, we obtain

$$\mathrm{d}S = \frac{V}{T}\frac{\mathrm{d}u}{\mathrm{d}T}\mathrm{d}T + \frac{4}{3}\frac{u}{T}\mathrm{d}V. \quad (\text{II}, 36)$$

Now dS is a pure differential in the sense of I (a "total differential") and can therefore be written:

$$\mathrm{d}S = \frac{\partial S}{\partial T}\mathrm{d}T + \frac{\partial S}{\partial V}\mathrm{d}V.$$

Consequently $\dfrac{\partial^2 S}{\partial T \partial V} = \dfrac{\partial^2 S}{\partial V \partial T}$

and hence, from (II, 36) we have

$$\frac{1}{T}\frac{\mathrm{d}u}{\mathrm{d}T} = \frac{4}{3}\frac{1}{T}\frac{\mathrm{d}u}{\mathrm{d}T} - \frac{4}{3}\frac{u}{T^2},$$

i.e. $\dfrac{\mathrm{d}u}{u} = 4\dfrac{\mathrm{d}T}{T}.$

Hence

$$u = a\,T^4, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{II}, 37)$$

where a is an integration constant. This is the Stefan-Boltzmann law. As the radiation energy, i.e. the quantity of radiation impinging per second upon 1 cm^2, is proportional to the quantity of energy per cm^3, this law may also be stated: the total radiation of a black body is proportional to T^4. Thus, substituting (II, 37) for u in (II, 36), we obtain:

$$\mathrm{d}S = 4\,aV\,T^2\,\mathrm{d}T + \frac{4}{3}a\,T^3\,\mathrm{d}V$$

so that $\left(\dfrac{\partial S}{\partial V}\right)_T = \dfrac{4}{3}\,a\,T^3. \quad . \quad . \quad . \quad (\text{II}, 38)$

The energy per unit volume of the radiation is thus proportional to the fourth power of the absolute temperature, and the entropy per unit volume to its cube.

ERRATA

THE THEORY AND CONSTRUCTION OF GERMANIUM DIODES

In the article by J. Z. van Wessem, The theory and construction of germanium diodes, which appeared in the February issue of this Review, the heading "The p and n concentrations in the absence of an applied voltage" on p. 218, should read: "The p and n concentrations in the *presence* of an applied voltage".

"MÜLLER" UGX APPARATUS FOR X-RAY DIAGNOSTICS

In the short notice concerning the above apparatus which appeared in our February issue (p. 237), it was stated that the patient can be rotated about an axis *parallel* to the X-ray beam. It should be pointed out that the axis of rotation is not merely parallel but *coincident* with the axis of the beam.